FUSEDAR: Adaptive Environment Lighting Reconstruction for Visually Coherent Mobile AR Rendering

Yiqin Zhao* Tian Guo†

Worcester Polytechnic Institute

ABSTRACT

Obtaining accurate omnidirectional environment lighting for high quality rendering in mobile augmented reality is challenging due to the practical limitation of mobile devices and the inherent spatial variance of lighting. In this paper, we present a novel adaptive environment lighting reconstruction method called FUSEDAR, which is designed from the outset to consider mobile characteristics, e.g., by exploiting mobile user natural behaviors of pointing the camera sensors perpendicular to the observation-rendering direction. Our initial evaluation shows that FUSEDAR achieves better rendering effects compared to using a recent deep learning-based AR lighting estimation system [8] and environment lighting captured by 360° cameras.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms Mixed / augmented reality—; Human-centered computing—Ubiquitous and mobile computing—Empirical studies in ubiquitous and mobile computing

1 INTRODUCTION

In this paper, we address the key challenges associated with reconstructing high-fidelity environment maps which can be subsequently used for rendering visually-coherent virtual objects with complex geometry and metallic material.

Via controlled experiments with a synthetic dataset, we develop a novel method called FUSEDAR, sharing similar spirit to the well-known screen space reflection [4], that transforms camera observation(s) to a hybrid environment map through mesh reconstruction and blurring/color filling. At a high level, FUSEDAR constructs mesh for the *near field*, portions that receive more accurate and higher confidence depth information surrounding rendering position. This design helps producing geometrically accurate lighting transformation between the observation and rendering positions, thus supporting key rendering features like reflections to provide visual continuity. Further, FUSEDAR leverages far-field observations to address the anisotropic property of lighting by reconstructing low-frequency lighting to reduce visual errors.

Our design achieves 36.7% and 17.1% higher rendering PSNR when comparing to a recent deep-learning based AR approach [8] and environment lighting captured by 360° cameras at the observation position, respectively.

2 DISTANCE-BASED LIGHTING RECONSTRUCTION

FUSEDAR is an adaptive pipeline that progressively, e.g., as AR users naturally move around the indoor environment, reconstructs environment lighting at a user-specified rendering position.

Near-field reconstruction. FUSEDAR leverages the increasingly popular depth sensor in mobile camera system [1,2] as depth sensors enable the possibility of capturing geometrically accurate environment observations. We design our near-field lighting reconstruction method based on mesh reconstruction. First, from each camera observation, we generate 3D point cloud using the RGB-D images and map the generated point cloud to AR world coordinates using the camera position and orientations. Then, to reduce the computation requirements, we perform voxel point cloud sampling on each generated point cloud with voxel size of 1 centimeter. The voxel size subsequently determines the reconstructed mesh precision. Next, we reconstruct meshes within a defined *near-field boundary*, which specifies the cubic length and corresponds to the size parameter of Poisson mesh reconstruction [3], to partition the environment observation into near and far-fields. Specifically, in this work, we set the near-field boundary to be 2.56 meters and points fall into the corresponding cubic space are considered as part of the near field while points outside belong to the far field. Finally, we trim the reconstructed mesh by removing vertices and faces outside of 99% input point cloud density space and then project the resulting mesh into a 360° environment map as part of the FUSEDAR reconstructed near-field lighting.

Far-field reconstruction. In addition to near-field mesh-based lighting reconstruction, FUSEDAR reconstructs far-field lighting by focusing on the anisotropic property of lighting to provide omnidirectional lighting as the reconstructed result. To reconstruct lighting in far-field, FUSEDAR first applies blurring on the far-field observation image using the normalized box filter [5] to remove high-frequency details on observation images. The blurring filter also smooths the image boundary to further increase the overall accuracy and visual continuity. Then, FUSEDAR copies the blurred far-field observation image to the corresponding position on the reconstructed environment map based on camera orientations. Next, for unobserved area on the generated environment map, FUSEDAR fills the empty pixels with the dominant color found from colors of existing observations. Specifically, FUSEDAR first combines observations from blurred far-field camera images and near-field reconstructed mesh to form an environment map. To address the anisotropic property of lighting, we leverage the unit-sphere point cloud sampling [8] to approximate unobserved far-field environment based on uniform sampling of existing direct observations. Concretely, FUSEDAR uses the unit-sphere point cloud sampling to sample the pixel/vertex color at directions given by a set of uniformly distributed points on a unit sphere to create a uniformly distributed observation color sample. Finally, we use the *KMeans* clustering method (we set $k = 1$) to find the dominant color.

3 EVALUATION

To evaluate FUSEDAR's effectiveness in rendering virtual objects in mobile AR, we generated a synthetic dataset using the Unreal Engine and photorealistic assets. The synthetic data includes ground truth lighting information, which is used as the basis for calculating Peak signal-to-noise ratio (PSNR) and Structural Similarity Index (SSIM). Specifically, FUSEDAR is compared to two baselines: *(i)* using a 360° degree camera at the observation position ; and *(ii)* a recent

---

*e-mail:yzhao11@wpi.edu

†e-mail:tian@wpi.edu

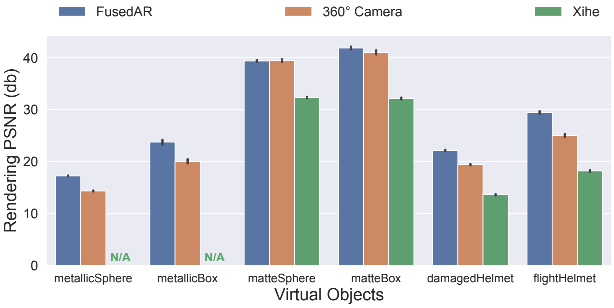

Figure 1: ***End-to-end rendering accuracy comparison.*** FusedAR achieves better rendering effects, i.e., higher PSNR values (calculated against ground truth rendering), than other techniques for all objects. The PSNR comparisons to Xihe are omitted for the first four virtual objects as Xihe only provides spatial-variant low-frequency lighting estimation [8].

academic framework Xihe that produces real-time low-frequency lighting estimation from RGB-D images [8]. We compare their visual performance both qualitatively and quantitatively (by calculating their respective PSNR and SSIM values against the virtual object rendered with ground truth lighting). We show that FusedAR outperforms both baselines for all six objects tested and achieves better visual similarity to ground truth rendering (§3.2).

3.1 Simulation Environment Setup

We first create a photorealistic indoor scene based on the *ArchViz* project [6], a high quality architectural visualizations for interior design and Unreal Engine. Using this scene, we create a virtual camera using the Blueprint programming system that takes controlled variables to modify the camera's movement and internal properties, e.g. FoV. In the created indoor scene, we first manually choose 10 reasonable positions to be considered as lighting reconstruction position for placing virtual objects. Example reconstruction positions include on the floor or table. For each reconstruction position, we vary a number of factors, including the number of capturing position, mobile user/device height, and observation distance, to generate 72 camera observations. For each camera observation, we export the camera HDR observation image, depth image, position, orientation, and ground truth lighting in the format of equirectangular panorama image. In total, our dataset consists of 720 camera observations.

We develop a browser-based renderer using the Three.js rendering framework [7] to automate the process of rendering virtual objects of interest. Specifically, our renderer uses information including reconstructed lighting, the camera position and properties, from our synthetic dataset to render a 3D virtual object at the resolution of 1024x768. The renderer then trims empty pixels outside rendered objects to remove the object-to-frame size impact on PSNR calculation when using different camera FoV settings. The resulting images of rendered objects serve as the basis for comparing different lighting reconstruction methods.

3.2 End-to-end Performance

We compare the end-to-end rendering performance, both quantitatively and visually, on six different virtual objects. For this experiment, we configure FusedAR to reconstruct the environment map with the following parameters: *(i)* one natural camera observation for near-field mesh reconstructed; *(ii)* nine more observations based on the guided movement for far-field reconstruction.

Figure 1 shows the rendering PSNR value comparisons. Specifically, FusedAR achieves 44.1% and 12.1% higher rendering PSNR than a recent deep learning-based AR lighting estimation system [8] and lighting captured by a 360° camera, on complex objects with physically-based materials (i.e., *damagedHelmet* and *flightHelmet*). This indicates that by correctly leveraging user movement and scene geometry information, FusedAR can generated highly accurate lighting from limited camera observations. Note that we omit the comparison to the rendering PSNR by Xihe on metallic objects (*metallicSphere* and *metallicBox*) as Xihe only provides low-frequency lighting in SH coefficients format, which does not support reflective rendering.

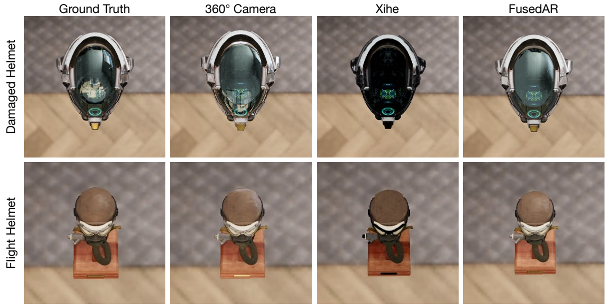

Figure 2: ***End-to-end rendering visual effect comparison.*** Virtual objects rendered with lighting provided by ground truth lighting, 360° camera, Xihe [8] and FusedAR.

Figure 2 compares the visual effect of objects rendered with different lighting information. We observe that FusedAR produces visually coherent virtual objects. This suggests FusedAR is effective in generating a completed high-fidelity environment map.

4 CONCLUSION

In this work, we introduced a novel adaptive pipeline FusedAR that generates completed high-fidelity environment maps from one or more HDR images and depth information. As demonstrated both quantitatively and qualitatively, our reconstructed environment maps can be used to render objects of various properties, including reflective materials, with better PSNR/SSIM and visual coherence than other tested techniques. FusedAR's distance lighting reconstruction can work with as few as one camera observation and can progressively improve the environment map quality, especially for metallic objects, with more camera observations.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grants #1815619 and #2105564, and VMWare.

REFERENCES

[1] HUAWEI. HUAWEI mate 30 pro specifications — HUAWEI global. `https://consumer.huawei.com/en/phones/mate30-pro/specs/`. Accessed: 2020-7-8.

[2] A. Inc. ipad pro 2020. `https://www.apple.com/ipad-pro/specs/`, 2020.

[3] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, vol. 7, 2006.

[4] M. McGuire and M. Mara. Efficient gpu screen-space ray tracing. *Journal of Computer Graphics Techniques (JCGT)*, 3(4):73–85, 2014.

[5] OpenCV. Image Filtering. `https://docs.opencv.org/4.5.3/d4/d86/group__imgproc__filter.html#ga8c45db9afe636703801b0b2e440fce37`. Accessed: 2021-11-5.

[6] Sciontidesign. Archviz Interior Rendering. `https://docs.unrealengine.com/4.27/en-US/Resources/Showcases/ArchVisInterior/`. Accessed: 2021-11-5.

[7] Three.js Organization. Three.js - JavaScript 3D library. `https://threejs.org`. Accessed: 2021-11-5.

[8] Y. Zhao and T. Guo. Xihe: A 3d vision-based lighting estimation framework for mobile augmented reality. In *The 19th ACM International Conference on Mobile Systems, Applications, and Services*, 2021.