Rendering for Hyper-Realistic Displays: A Benchmark Study

Akshay Jindal
https://www.cl.cam.ac.uk/~aj577
Rafał Mantiuk
https://www.cl.cam.ac.uk/~rkm38
Department of Computer Science and Technology,
University of Cambridge

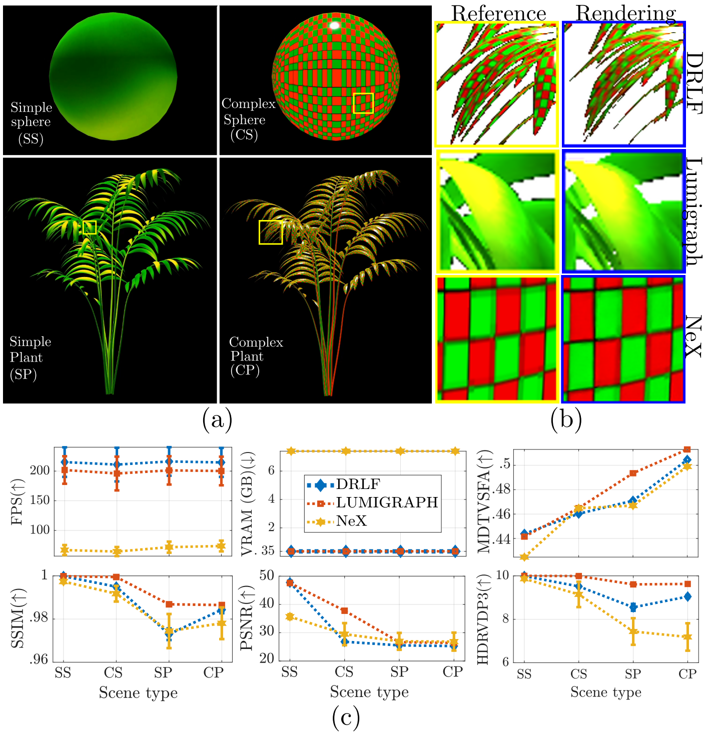

Figure 1: 3D scenes used for comparing different IBR methods (a), sample artefacts caused by three rendering methods (b), and performance and qualitative comparison of the three rendering methods (c).

Introduction

Hyper-realistic displays [5] are an emerging class of computational displays that strive to reproduce reality by delivering a large subset of perceptual realism cues such as high spatio-temporal resolution, high dynamic range (HDR), binocular disparity, motion parallax, accommodation, and colour. The rendering methods for these displays are required to process high-resolution HDR images in real-time to generate novel views from arbitrary viewpoints while delivering correct depth cues. These requirements make image-based rendering (IBR) methods a natural choice for such displays due to their relatively low computational cost and high-quality results. Each method comes with its own set of trade-offs; thus, it is necessary to understand how their artefacts translate to perceived quality. In this work, we compare three real-time techniques from different ends of the IBR spectrum: dynamically reparameterised lightfield (DRLF) [2], lumigraph implemented as a mesh with view-dependent textures [5], and NeX, a neural multi-plane images method [4]. We present performance benchmarks for the three algorithms, rate their visual artefacts using objective quality metrics, and show how the existing quality metrics are insufficient to evaluate the quality of IBR techniques.

Method

We compare the three methods on 4 forward-facing synthetic scenes (2 geometries × 2 materials) generated using Unity3D (Figure 1a). The geometries used were a sphere mesh (2 800 vertices and 960 triangles) and a plant mesh (254 244 vertices and 84 748 triangles). The meshes were mapped to two materials: a Lambertian material with a low-frequency gradient image as its diffuse component and a specular material (Phong shading) with a high-frequency checkerboard as its diffuse component.

We rendered 20 images of 2160×1440 resolution and 16-bit per channel with a baseline of 100 mm for all 4 scenes. Even numbered images were fed to the rendering methods for novel view synthesis, and the odd images were used for validation. The three methods were implemented in MATLAB and OpenGL and were used to generate new views from the perspective of validation images to facilitate objective quality evaluation. The computational performance of each method was measured over 1000 frames on an i7-8700 CPU @ 3.20GHz, 32GB RAM and an RTX 2080Ti GPU with V-Sync disabled. In addition to the validation poses, 60 Hz videos were rendered and are available on the project web page[1].

Results

The DRLF method is computationally the simplest method with the highest average FPS and lowest VRAM requirements (Figure 1c). The lumigraph method is a close second in terms of performance and memory requirements. The memory required by its mesh representation is negligible compared to HDR images. There is also no noticeable effect of the number of triangles between sphere and plant scene, indicating that mesh size is not a bottleneck on modern GPUs. NeX is the slowest and most-memory intensive method (20× more memory). We found one of the biggest bottlenecks in NeX's performance to be the texture-lookup operation on MPI LUTs. An implementation that makes the better use of the GPU memory architecture might improve performance.

The artefacts induced by the three methods are also quite different (Figure 1b). DRLF works well on the simple sphere but blurs the regions away from the focal plane in other scenes. This is visible as blurred textures or missing thin edges. Lumigraph renders sharp textures in all scenes but maps incorrect textures near thin edges (disocclusions). Since lumigraph uses the rasterization of a mesh, the geometry edges can also suffer from aliasing artefacts. In both methods, these artefacts become more prominent in videos and appear as distracting flickering near thin edges or juddery specular highlight. NeX works well on both thin edges and specular highlights but gravely affects the texture appearance of checkerboard material. Also, NeX induces unnatural halos around the object when the rendering viewpoint is far from MPI's reference pose.

To quantify the quality of each method, we compared their renderings for the validation camera views against the validation images using 3 popular objective quality metrics: PSNR, SSIM, and HDR-VDP-3. Since PSNR and SSIM were designed for SDR images, the HDR images were first encoded using PU21 transform [1]. HDR-VDP-3 assumed a linear RGB colour space and 45 ppd resolution. The plots in (Figure 1c) show the quality results averaged over 10 validation images, and the error bars show the standard deviation. We also run a blind video quality assessment metric MDTVSFA [3] on the videos recorded for each method. Overall, the lumigraph rendering method is consistently rated as the highest quality in all 4 scenes. However, there is not much consensus on the rating and ranking of the methods across 4 metrics. Lumigraph and DRLF have similar quality (low std. deviation) across all 10 validation images (except DRLF for SP), but NeX has a higher variance due to artefacts in images far from the reference pose. Since the metrics we used are some of the most popularly used metrics in IBR evaluation, it is crucial to understand which of these metrics best relate to human ratings. Note that none of these metrics were designed for IBR artefacts, and most of them do not account for HDR or temporal artefacts such as judder and flicker.

[1] M. Azimi and R. K. Mantiuk. Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr. In *2021 Picture Coding Symposium (PCS)*. IEEE, 2021.

[2] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000.

[3] D. Li, T. Jiang, and M. Jiang. Unified quality assessment of in-the-wild videos with mixed datasets training. *International Journal of Computer Vision*, 2021.

[4] S. Wizadwongsa, P. Phongthawee, J. Yenphraphai, and S. Suwajanakorn. Nex: Real-time view synthesis with neural basis expansion. In *Proc. of Computer Vision and Pattern Recognition*, 2021.

[5] F. Zhong, A. Jindal, A.O. Yöntem, P. Hanji, S. J. Watt, and R. K. Mantiuk. Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Trans. Graph.*, 2021.

[1] `https://www.cl.cam.ac.uk/research/rainbow/projects/hdrmfs/rendering_methods/`