Reflectance Estimation for Free-viewpoint Video

George Ash
Dimension
United Kingdom
george@dimensionstudio.co

Juraj Tomori
Dimension
United Kingdom
juraj@dimensionstudio.co

Mike Pelton
Dimension
United Kingdom
mike@dimensionstudio.co

Charles Dupont
Dimension
United Kingdom
charles@dimensionstudio.co

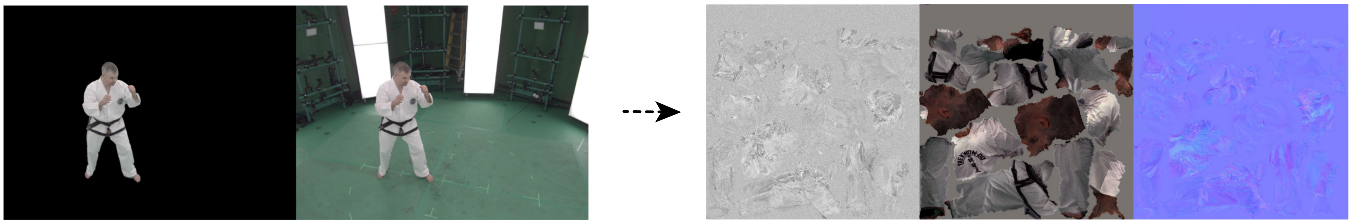

Figure 1: Left: renderer output, ground truth image feed. Right: the output of our method: roughness, albedo, normal maps.

ABSTRACT

We present a method to infer physically-based material properties for free-viewpoint video. Given a multi-camera image feed and reconstructed geometry, our method infers material properties, such as albedo, surface normal, metallic and roughness maps. We use a physically based, differentiable renderer to generate candidate images which are compared against the image feed. Our method searches for material textures which minimise an image-space loss metric between candidate renders and the ground truth image feed. Our method produces results that approximate state of the art reflectance capture, and produces texture maps that are compatible with common real-time and offline shading models.

ACM Reference Format:
George Ash, Juraj Tomori, Mike Pelton, and Charles Dupont. 2021. Reflectance Estimation for Free-viewpoint Video. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '21 Posters), August 09-13, 2021.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3450618.3469146

1 INTRODUCTION

Capturing photorealistic, high quality performances of humans for use in film and video games remains an active area of research in computer graphics and computer vision. Whilst there have been significant developments in reflectance capture for free-viewpoint video [Guo et al. 2019], less work has been done to augment captures under a single lighting condition, which is most commonly in use for free-viewpoint video captures today.

Previous work in free-viewpoint video [Collet et al. 2015] captures geometry and a diffuse texture, but illumination is baked onto the surface of the capture, and the method struggles to reconstruct specular surfaces. Later work [Guo et al. 2019] improves relightability by computing albedo, surface normal, shininess and ambient occlusion maps, but requires a complex lighting setup producing spherical gradients and careful synchronization between lighting changes and the video feed. Our work assumes the same 106-camera set up as in [Collet et al. 2015], but improves on the method, by separating the 'baked in' diffuse texture into several physically-based maps, allowing for improved relightability and reproduction of view-dependent reflectance.

By framing the problem as an inverse-graphics optimisation procedure: that is, minimising the image-space loss between candidate renders and a ground-truth image feed from the camera array, we can optimise for the following material properties: albedo, roughness, metalness and normals.

2 OUR METHOD

To allow our candidate renders to match the camera feed most closely, we implemented a differentiable physically-based shading model in PyTorch, which uses nvdiffrast [Laine et al. 2020] for fast, differentiable rendering abstractions.

2.1 Lighting

We implement a GGX microfacet BRDF, described in [Karis 2013]. We can then accurately light our base geometry under a prefiltered HDR environment map of the capture stage (seen in figure 2), and the material texture variables. We found the need to inhibit strong specular effects in cloth areas, so we multiply the specular term by a weighting factor, which we set up as a variable texture map in the optimisation procedure.

2.2 Image Loss

To ensure our candidate renders and camera feed line up in image space, we use the same callibration procedure described in [Collet et al. 2015]. We can then combine the camera extrinsics, such as world space position and rotation with the camera intrinsics, such as focal length to produce a pinhole projection matrix. We also

SIGGRAPH '21 Posters, August 09-13, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8371-4/21/08.
https://doi.org/10.1145/3450618.3469146

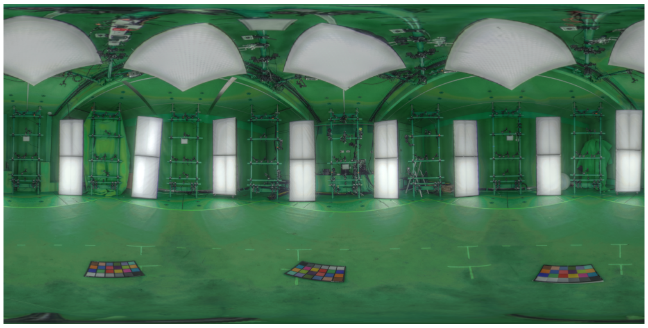

Figure 2: The HDR environment map of the capture stage

Figure 3: Comparison in reconstructing higly specular materials. Left: our method. Right: Collet et al.

apply an inverse distortion function to the camera feed to correct for barrel distortion introduced by the camera lens.

Since our rendered images are computed in a high dynamic range colourspace, they need to be tone-mapped before comparison with the camera feed. We apply a tone curve [Reinhard et al. 2002] on the rendered image before computing the root-mean-square error against the undistorted ground truth image.

2.3 Optimisation procedure

For each step in the optimisation procedure, we choose a camera uniformly from the multi-camera array. We then render a candidate image using the camera parameters. We take advantage of automatic differentiation offered by PyTorch to compute gradients of the loss function with respect to our texture map variables. We use Adam optimiser [Kingma and Ba 2014] to search the parameter space for textures that minimize the image loss.

Texture variables are initialised to a resolution of 2048x2048. We run the optimisation procedure for 2000 iterations over approximately 100 minutes on an NVIDIA A6000 GPU.

3 RESULTS

Our work shows a visible improvement over [Collet et al. 2015], particularly for highly specular areas such as the jacket in figure 3.

Our method produces comparable textures to the method by [Guo et al. 2019], but it does not require an alternating illumination gradient at capture time.

3.1 Limitations

Unlike [Guo et al. 2019], our method does not consider ambient occlusion or shadowing when calculating reflectance at a point on the surface, and some lighting fluctuations can leak into the albedo texture. Other view-dependent effects may also leak into the albedo texture where there is poor coverage from the camera array, and the optimisation problem is under-constrained under a single lighting condition.

3.2 Future Work

We would like to address the limitations of our method with a learned model and regularisation terms that can help infer plausible texture maps, resolving ambiguities in the optimisation procedure.

REFERENCES

Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-Quality Streamable Free-Viewpoint Video. *ACM Trans. Graph.* 34, 4, Article 69 (July 2015), 13 pages. https://doi.org/10.1145/2766945

Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, Danhang Tang, Anastasia Tkach, Adarsh Kowdle, Emily Cooper, Mingsong Dou, Sean Fanello, Graham Fyffe, Christoph Rhemann, Jonathan Taylor, Paul Debevec, and Shahram Izadi. 2019. The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting. *ACM Trans. Graph.* 38, 6, Article 217 (Nov. 2019), 19 pages. https://doi.org/10.1145/3355089.3356571

Brian Karis. 2013. Real Shading in Unreal Engine 4. In *ACM SIGGRAPH 2013 Courses: Physically Based Shading in Theory and Practice* (Anaheim, California) *(SIGGRAPH '13)*. Association for Computing Machinery, New York, NY, USA, Article 22, 8 pages. https://doi.org/10.1145/2504435.2504457

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. 2020. Modular Primitives for High-Performance Differentiable Rendering. *ACM Trans. Graph.* 39, 6, Article 194 (Nov. 2020), 14 pages. https://doi.org/10.1145/3414685.3417861

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic Tone Reproduction for Digital Images. *ACM Trans. Graph.* 21, 3 (July 2002), 267–276. https://doi.org/10.1145/566654.566575