CSC2529H Direct and Indirect light NeRF

Frank Sun and Anagh Malik

November 2022

1 Motivation

When we take an image of the scene, by summing up the photons over the exposure time we discard the information of their arrival and paths taken. Emerging sensors have, however, been spearheading a new imaging technique which allows you to view a trillion frames per second. This can even facilitiate viewing single photon arrivals. In the activate imaging regime, transient imaging might allow disambiguating between the the direct light (which contains photons which have bounced off at most 1 surface) and indirect light (which has bounced off at least 2 surfaces). By disambiguating betweens these photons, we can even create images of direct light and indirect light. Capturing this image data is however expensive. Recent developments in 3D Computer Vision have however led to algorithms for Novel View Synthesis, which would allow us to recover indirect/direct light images from just a few multiview observations.

2 Related Works

Scene Reconstruction. There exists a wide range of methods for RGB and RGB-D based 3D reconstruction. Most of RGB-D reliant methods are based on [1], where multiple depth measurements are fused using a signed distance function (SDF) which is stored in a uniform 3D grid. An example of such work is KinectFusion [6] combines such representation with real-time tracking to reconstruct objects and small scenes in real-time. An example of a method reliant on just RGB images is Single View MPI [9], which learns to generate multiplane images given one or more images with known viewpoints.
Most recently coordinate-based multi layer perceptrons (MLP) have become a popular representation of the 3D scene [3]. As input the MLP takes a 3D location in the model space and outputs for example, occupancy, density or colour. There has been a lot of work using this simple idea in multiple applications like SLAM [8], [11], for novel-view synthesis [3], [10].
Direct/Indirect Light Separation One of the papers that sparked the interest in direct and indirect separation was published in 2006 [5]. The papers proposes to use high frequency light sources and multiview images to get light

separation. Since then there has been a lot of follow up work [4], [7], however none have attempted to reconstruct novel indirect/direct light views.

3 Project Overview

For the project we will be working in simulation, specifically we use the mitsuba renderer [2] to render transient images of the scene. We will then use a heuristic to separate direct/indirect photons, which can then be added to up to give the indifrect and direct component images. We will then use Nerf [3] to be able to render novel indirect and direct images.

4 Milestone, timeline and goals

There are many fast NeRF implementations, which make the novel view synthesis part easier and might take only a few days to get the dataloader and train some good looking scenes. It will be more time consuming to find interesting scenes and render them out with mitsuba – finding camera directions etc might take some time and use of blender. All in all the project should take at most 3 weeks.

References

[1] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, page 303–312, New York, NY, USA, 1996. Association for Computing Machinery.

[2] W. Jakob, S. Speierer, N. Roussel, M. Nimier-David, D. Vicini, T. Zeltner, B. Nicolet, M. Crespo, V. Leroy, and Z. Zhang. Mitsuba 3 renderer, 2022. https://mitsuba-renderer.org.

[3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[4] O. Nasu, S. Hiura, and K. Sato. Analysis of light transport based on the separation of direct and indirect components. 06 2007.

[5] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, 25(3):935–944, jul 2006.

[6] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion:

Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011.

[7] M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *ACM Trans. Graph.*, 33(4), jul 2014.

[8] E. Sucar, S. Liu, J. Ortiz, and A. J. Davison. imap: Implicit mapping and positioning in real-time. *CoRR*, abs/2103.12352, 2021.

[9] R. Tucker and N. Snavely. Single-view view synthesis with multiplane images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[10] D. Verbin, P. Hedman, B. Mildenhall, T. E. Zickler, J. T. Barron, and P. P. Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. *CoRR*, abs/2112.03907, 2021.

[11] Z. Zhu, S. Peng, V. Larsson, W. Xu, H. Bao, Z. Cui, M. R. Oswald, and M. Pollefeys. NICE-SLAM: neural implicit scalable encoding for SLAM. *CoRR*, abs/2112.12130, 2021.