Project Proposal 16-825:

Non Line of Sight Rendering (N-LoSeR)

Udit Patadia (uap@andrew.cmu.edu)
Akshay Venkatesh (akshayv@andrew.cmu.edu)

Outline & Motivation

Our project aims to come up with a model that can generate rendered images of an object that is occluded (hidden behind a corner of a wall).

In essence, what we are trying to do is to learn the parameters of diffuse reflection along the light path to the object being observed so that we can perform a de-blurring on the reflection seen on the diffuse surface to recover the object of interest. (tldr: camera looks at wall, recovers reflection of object not in field of view)

We also want to evaluate whether our network accurately learns the scene parameters and whether the same network will work for recovering the images of other objects of interest(not used in training, generalizability), under different light conditions(like with ambient light or environment light).

We expect our network to learn the parameters for a scene with just a single change in object of interest as it may be sufficient for cases like surveillance using a fixed camera looking around blind-spots where the scene would mostly be constant but there may be a few single/ stray objects that are different, which would hopefully not alter the lighting conditions a great deal. An evaluation of how robust our approach is to lighting conditions and changes is also a part of our plan.

Another aspect of our work would be to see how well this translates to simple real-world situations. The motivation being that if we can train the network on simulated scenes and we can accurately estimate the reflectances of the surfaces, does that help us generate a model that can be deployed in the wild successfully.

A lot of our motivation is derived from the reading material listed in the references section which include both classical and learning based approaches to non line-of-sight imaging and computational periscopy.

Approach

Our network will take in a camera location, viewing direction as input, and an image of a diffuse surface as input. The output would be a single image render, of hopefully the occluded object. The ideal render for each case would be the scene if the surface being viewed were a perfect mirror. For a start, we would like to work with a NeRF[1] like architecture that also takes in an image in addition to the position and direction. The hope is that the rendering parameters for that viewing direction (transmittances and densities) would be subsumed into the network upon training for multiple views. The architecture and understanding would likely require more refining upon training and evaluating what the network is learning and we would probably use inspiration from [2] as well to help motivate our network architecture..

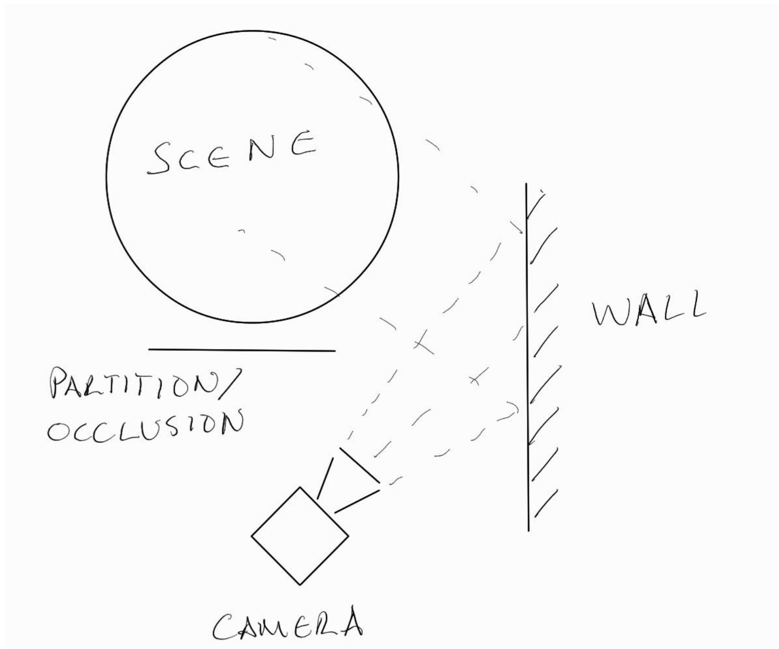

Figure 1: Layout illustration

Training

We hope to construct a scene as shown in the image in simulation like Blender, and train the network with ground truth images and camera poses collected for an array of camera locations, where the camera remains on the same side as the occluding partition and continues to look at the diffuse surface (wall) from which the reflection of the scene to be observed is un-diffused. The ground truth images will simply be the images rendered from the camera by setting the reflectance of the wall being viewed to that of a perfect mirror in our blender scene. The loss metric would be the mean squared error between the RGB image output by the network and the provided ground truth image for that viewing direction. Based on the performance of our system we plan to iterate over the scene settings (lighting conditions, wall reflectance etc.). Similar to NeRF we will try positional encoding to hopefully capture high frequency detail.

Test/ Eval/ End Goal:

We hope to evaluate the performance of our approach on three fronts:

- ability to generate renders of the scene for any generated camera pose on the same side of partition
- ability to render updated scene without having to train the network again
- ability to translate to real-world conditions (**stretch goal**)

References

[1] NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
[2] Non-line-of-Sight Imaging via Neural Transient Fields
[3] Physics to the Rescue: Deep Non-line-of-sight Reconstruction for High-speed Imaging
[4] Deep Non-Line-of-Sight Reconstruction
[5] Recent Advances on Non-Line-of-Sight Imaging: Conventional Physical Models, Deep Learning, and New Scenes
[6] Computational periscopy with an ordinary digital camera
[7] Diffuse Mirrors: 3D Reconstruction from Diffuse Indirect Illumination Using Inexpensive Time-of-Flight Sensors