Near-field photometric stereo with point light sources

Rakshith Srinivasa Murthy
rakshits@andrew.cmu.edu

ABSTRACT

In this project, I aim to extract the 3D shape and depth of objects with the "Near Field Photometric Stereo (NFPS)" problem setup. Using NFPS in practical scenarios is much more easier and convenient especially when dealing with large objects in constrained spaces. This is a more difficult problem than the regular far field approximation due to the changes in the underlying imaging model. In this report, I first highlight the fundamental differences between the "Far Field Photometric Stereo (FFPS)" and NFPS. In NFPS, the image intensity is non linear with respect to the depth of a point and therefore it is not possible to model it with linear equations. We will be using a neural network based approach in this report to tackle the problem of NFPS. We "recursively" update the normals and depths at multiple scales as opposed to just solving a classical optimization problem. The use of such a "recursive" algorithm helps in two ways - it makes the solution very fast and efficient, and such a recursive approach also helps in dealing with outliers while predicting the final depth. We consider the case of uncalibrated light sources which are more suitable for practical scenarios. In this report, I include the results of using this algorithm on the LUCES dataset and also images taken by me using the near field setup.

INTRODUCTION

The task of shape extraction from photometric stereo (PS) is a well established one. In this task we model the surface normals, albedos and illumination as unknowns and make some assumptions on the type of surface (eg: Lambertian) to estimate the shape using a simple linear system. In general, photometric stereo is modelled with distant and directional light sources, which makes the underlying reflection equation quite simple. But if we consider nearby point light sources (near field approximation), the problem is more complicated and is no longer linear. In this project I wish to extend the simple setup of photometric stereo for nearby point light sources.

In the near-field setting, the image intensity depends non-linearly on the 3D location and normal of the scene point as well as the 3D light source position.

For objects close to a perspective camera, the assumption of orthographic projection model also fails. Therefore, the resulting problem is basically the optimization of a non linear and a highly non convex function where the initialization is very important.

Practically speaking, near field photometric stereo allows for more accurate reconstruction of the 3D surface of an object compared to traditional photometric stereo methods that assume distant light sources. By capturing images with light sources in close proximity to the object, subtle surface details and fine-scale features can be better recovered.

I firstly start off by explaining the problem setup of PS in the far field case. Throughout this report, I only limit to objects with the surfaces which obey Lambertian reflectance properties. Therefore using the far field approximation we get following imaging model.

$$I^i(p(X)) = \rho(X) n^T(X) d^i \tag{1}$$

Where X is the 3D point under consideration, I^i is the grayscale intensity image produced with the i^{th} light source. $p(X)$ is the location of that point in the pixel coordinates of the image. $\rho(X)$ is the albedo value of the surface at location X. d^i is the unit 3D vector representing the direction of the light from the i^{th} light source and $n(X)$ is the surface normal of the object at the 3D point X. Also the light sources here are assumed to have unit intensity.

The above image formation model is specifically for the case of far field approximation with the orthographic camera assumption.

But if we consider nearby point light sources, the value of the light intensity is no longer constant throughout the scene but rather it drops off according to the inverse square law.

Again let us consider a light source with unit intensity which is located at the 3D location L^i, then the new image formation model would look something as follows:

$$I^i(p(X)) = \rho(X) n^T(X) \frac{L^i - X}{\|L^i - X\|^3} \tag{2}$$

We have to note that since we are using a perspective camera the pixel location $p(X)$, depends on the depth $D(X)$ of a particular point X. Therefore, if we know the intrinsics K of the camera that we are using, we can estimate the 3D point X corresponding to the pixel $p(X)$ as follows:

$$X = D(X) K^{-1} p(X) \tag{3}$$

Therefore by using equation (3) in equation (2), we get the final image formation model as follow:

$$I^i(p(X)) = \rho(X) n^T(X) \frac{L^i - (D(X) K^{-1} p(X))}{\|L^i - (D(X) K^{-1} p(X))\|^3} \tag{4}$$

Now here we see that the grayscale intensity on a pixel $p(X)$ depends on the depth at that point in a non linear way. This is what makes NFPS an inherently difficult problem to solve. There have been multiple approaches to solve this which use both classical optimization based methods as well as modern deep learning based solutions.

1 PRIOR WORK

There have been quite a few papers trying to solve the task of depth and 3D reconstruction of objects in the case of Near Field Photometric Stereo. The authors in [8] propose a solution for the problem of uncalibrated NFPS which they solve with an alternating minimization scheme by first fixing the mean depth of the scene.

The authors in [9] however make use of a numerical solution to implement the optimization and estimate the depth. But they limit their approach to highly accurate calibrated light sources.

The algorithm used in [6] is another classical optimization method to estimate accurate 3D shape of objects in the NFPS setup. The

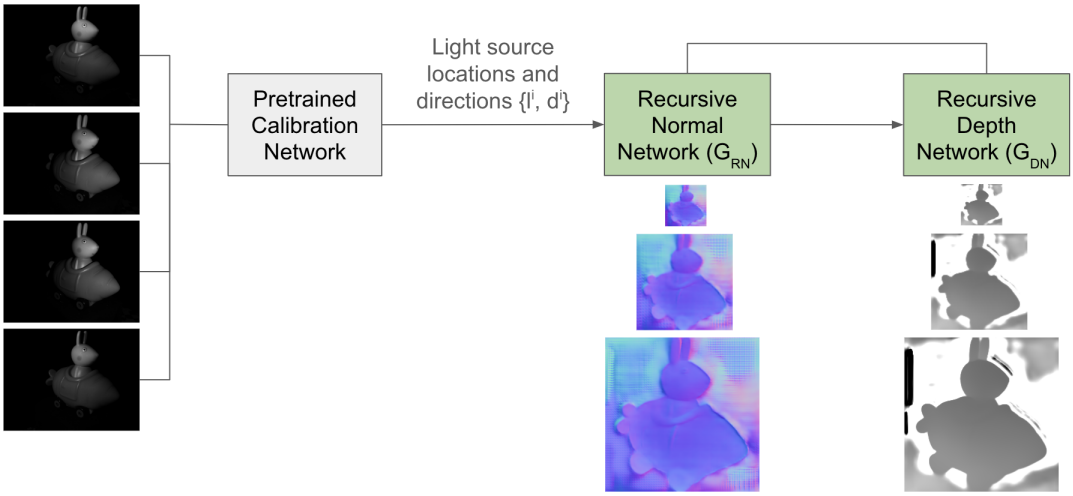

Figure 1: Overview of the recursive architecture.

authors here make use of a circular light array for all their experiments. They also require higly calibrated light sources. Also due to the inherent ambiguity and the non convex nature of the objective function involved, they highlight the importance of using a differential image for estimating a good initial guess for the depth estimate. The use of a circular light array on a plane perpendicular to the optical axis is really crucial for analytically estimating this initial guess. Therefore this approach is too restrictive to be applicable in most practical scenarios.

There are also statistical approaches as shown in [1] which tries to estimate the depth and 3D shape from just a single image. But to do this the authors of the paper use a lot of statistical priors to estimate the most plausible shape given the single input image.

On the other hand there are other learning based solutions such as [7] which make use of convolutional neural networks to directly predict the per pixel normals for the image and then integrate this normal field with the Poisson equation to get the resulting depth. But this paper also requires highly accurate calibrated light sources.

Finally the authors in [5] and [4] utilize a recursive algorithm using CNNs operating on the underlying image at multiple scales. This not only makes it more efficient than all the other approaches (including the optimization based approaches) but also more accurate to outliers for the final depth prediction.

The method used in this report is mostly based on this algorithm. Also at the end I also include results of using the final normal field and an additional mask to compute the depth map through Poisson integration.

METHOD

In this section, I provide a high level explanation for the depth reconstruction from a set of M images $I^1, I^2, \ldots, I^M$ of the same static scene in the presence of M different locations of a nearby point light source.

Also throughout this method, we assume that we know the intrinsics K of the camera that we use to capture the images. The authors in [4], use two separate recursive neural networks G_{RN} and G_{RD} to predict the normal and depth map respectively for an input image at a given scale. These networks are recursively called by increasing the resolution of the image. Since this network is pretrained on a large synthetic dataset, in this report I use the pretrained networks released by the authors. The use of such a recursive solution allows the depth prediction network to look at the entire image at a particular scale. Such a recursive architecture to predict normals and depth was first introduced in [5].

Initially the resolution of the images is 64x64, this is successively upsampled to finally reach the original resolution of the image. Along with the input images, we also use a per-pixel lighting image and attenuation. For a given depth image D, intrinsic K and the light source parameters (attenuation (μ), location (l) and direction (d)), we can compute it as follows for a pixel location (u, v):

$$X[u,v] = D[u,v]K^{-1}(u,v,1)^T \tag{5}$$

$$L[u,v] = normalize(X[u,v] - l) \tag{6}$$

$$A[u,v] = \frac{(L[u,v] \cdot d)^{\mu}}{\|X[u,v] - l\|^2} \tag{7}$$

These per pixel lighting and attenuation images will be utilized by the normal and depth prediction networks along with the normal and depth estimates from the previous scale to estimate the new normals and the depth at the current scale. The initial per pixel lighting images (L_0, A_0) are computed assuming the depth image D_0 is at unit distance.

A brief explanation of this pipeline is specified in Figure 1 with the outputs shown at a few different scales.

Also here we are dealing with uncalibrated light sources, where we dont know the exact locations and directions of the point light sources used. For this we use a pretrained network provided by authors in [2]. This network takes as input the set of M images $I^1, I^2, \ldots, I^M$ and predicts the positions $l^1, l^2, \ldots, l^M$ corresponding to each of the point light sources.

Finally, in this report I also include a method which can integrate the full resolution normals which are provided by the neural network along with a binary mask of the object to specify the boundary conditions. This will provide us with another estimate of the depth map limited to the extent of the object. This is not a substitute for the recursive depth prediction network G_{RD}, because the output from this network in the intermediate steps is necessary for computing the normals and the per pixel lighting images.

The results for using just the neural network for predicting both the normals and the depth as well as integrating the depths from the predicted normal field is shown in the next section.

RESULTS

In this section we show results on images from the LUCES dataset as well as images captured by me with point light source and perspective camera assumption. LUCES is the first real-world "dataset for nearfieLd point light soUrCe photomEtric Stereo" of 14 objects of different materials. 52 LEDs have been used to light each object positioned 10 to 30 centimeters away from the camera. The images captured by me are also captured with point light sources at 12 separate locations. For all the experiments and results, I used grayscale images of a static scene captured from the same viewpoint with varying light sources.

The normal, depth and 3D reconstruction for the LUCES dataset along with the ground truth visualization is depicted in Figure 2. As we can see that the normal and depth prediction in Figure 2 is not very good in some cases as shown in the first row. This could be due to the incorrect estimation of the intrinsic matrix for the camera used to capture these images. It could also be attributed to the fact that this network is trained originally on synthetic data and does not transfer well to real world images.

In Figure 3 I have added results for the images captured by me to estimate the surface normals, depth and finally reconstruct the 3D shape.

In Figures 4 and 5 I include the results to showcase the depth maps obtained by integrating the final predicted normal fields and a binary object mask to determine the boundary conditions. The binary object mask was obtained by using a pretrained "Segment Anything Model (SAM)" [3].

CONCLUSION

In this project, I explored different methods to estimate the surface normals and depth of objects to eventually reconstruct their 3D shape in the near field photometric stereo setup. Even though more convenient and practically applicable, the problem of NFPS is a much harder one as compared to the far field approximation. There are two main types of algorithms which aim to solve this namely classical optimization methods and deep learning based solutions. Based on my research I noticed that we need to have multiple assumptions and constraints on our system to analytically or numerically solve an optimization algorithm to estimate the depth and surface normals. Also most approaches require a highly calibrated setup to capture images in the presence of point light sources which are not feasible in most practical applications. But deep learning based solutions on the other hand have lesser number of constraints and are applicable to more scenarios. In this project I used one such approach of estimating the surface normals and 2D depth map directly by using a CNN recursively. I also tried another approach of using the final predicted normals to integrate and obtain the 2D depth map. Currently I utilized a pretrained network to estimate the depth and the results can be improved. As a future work, we can try to make the network scale better to real world data and also extend the neural networks to model more general materials other than just Lambertian surfaces.

REFERENCES

[1] Jonathan T Barron and Jitendra Malik. 2014. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence* 37, 8 (2014), 1670–1687.

[2] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. 2019. Self-calibrating deep photometric stereo networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8739–8747.

[3] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643* (2023).

[4] Daniel Lichy, Soumyadip Sengupta, and David W Jacobs. 2022. Fast light-weight near-field photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12612–12621.

[5] Daniel Lichy, Jiaye Wu, Soumyadip Sengupta, and David W Jacobs. 2021. Shape and material capture at home. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6123–6133.

[6] Chao Liu, Srinivasa G Narasimhan, and Artur W Dubrawski. 2018. Near-light photometric stereo using circularly placed point light sources. In *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–10.

[7] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. 2020. A cnn based approach for the near-field photometric stereo problem. *arXiv preprint arXiv:2009.05792* (2020).

[8] Thoma Papadhimitri and Paolo Favaro. 2014. Uncalibrated near-light photometric stereo. (2014).

[9] Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, François Lauze, and Jean-Denis Durou. 2018. Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision* 60 (2018), 313–340.

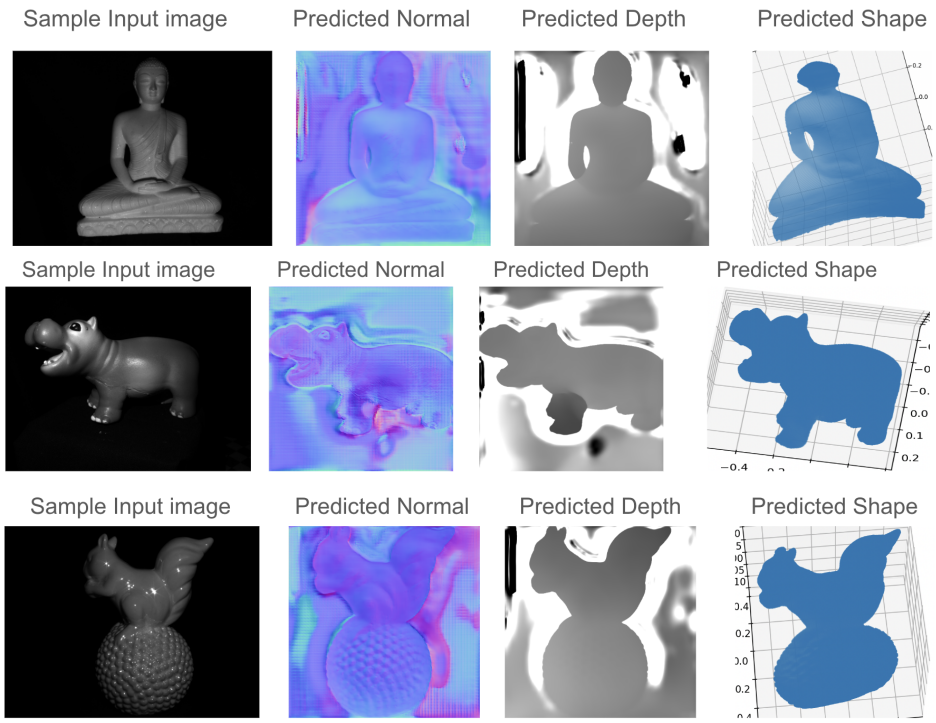

Figure 2: Normal and Depth prediction on samples from the LUCES dataset

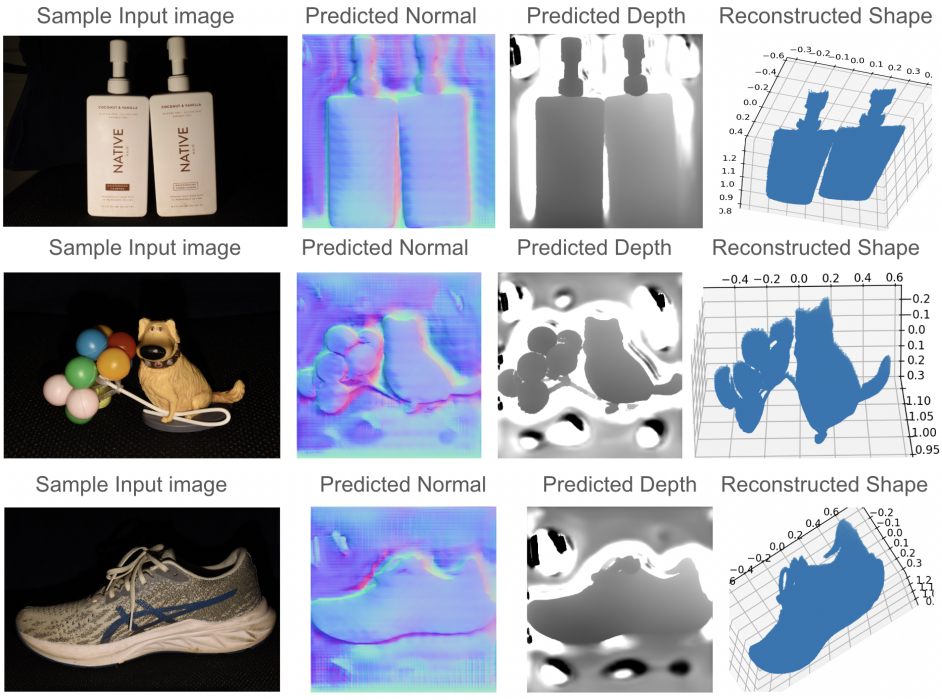

Figure 3: Normal and Depth prediction on images from my set up

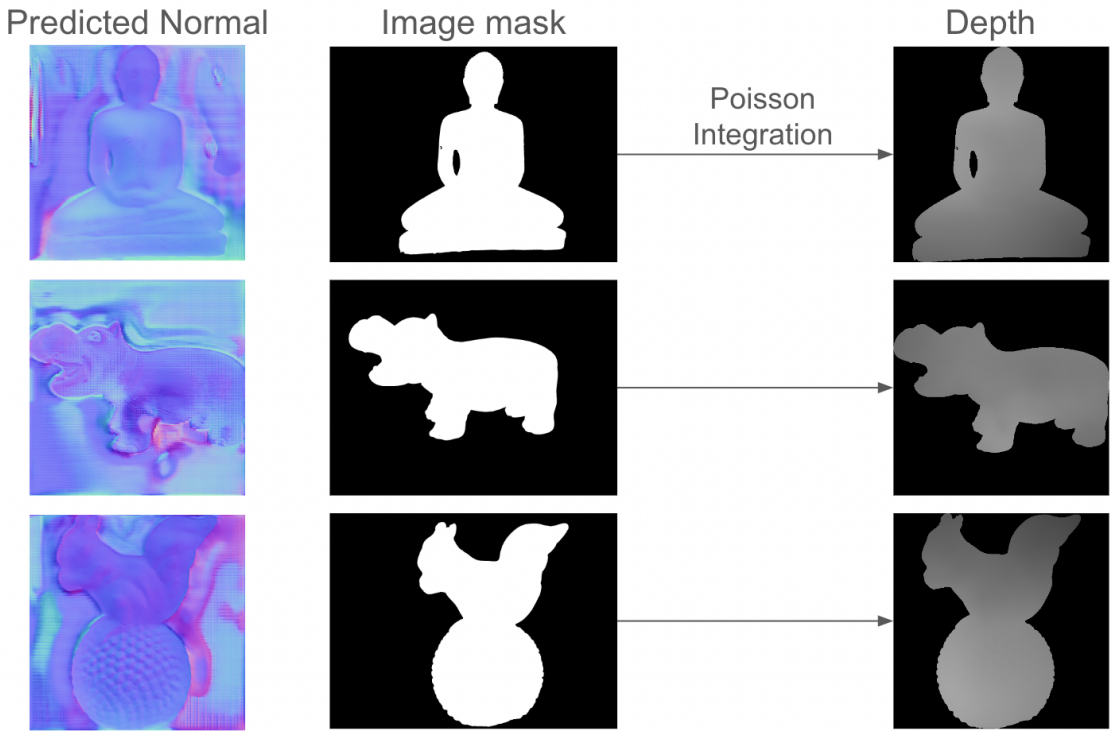

Figure 4: Depth by integrating the predicted normals on the LUCES dataset

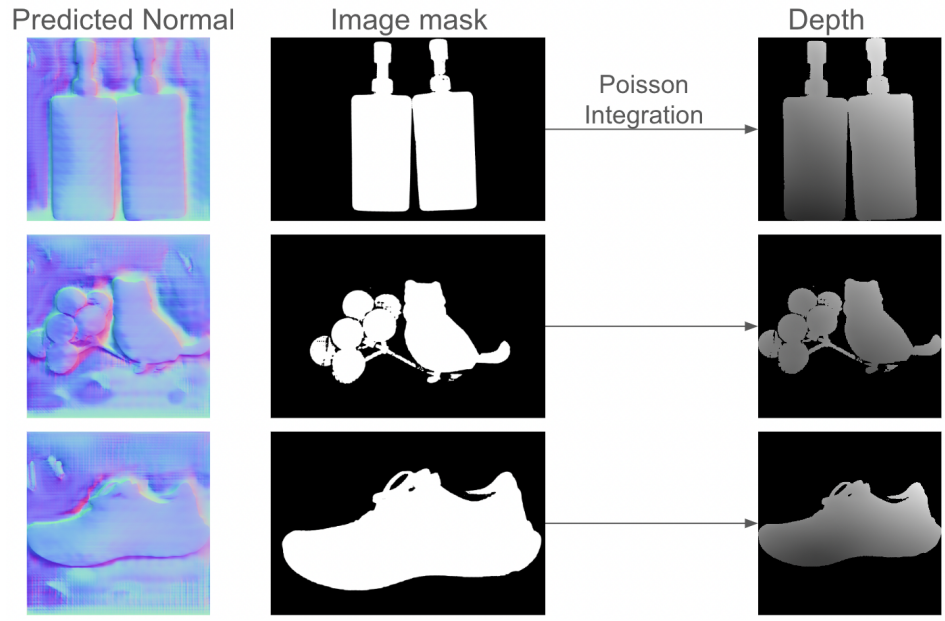

Figure 5: Depth by integrating the predicted normals on images from set up