A Neural Reconstruction Method for Non-Line-of-Sight Imaging

Salvador Rodriguez-Sanz, Diego Gutierrez, Albert Redo-Sanchez

Affiliation: Graphics & Imaging Lab (GILab, T34_23R)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: salvador.rodriguez@unizar.es

Abstract

Non-Line-of-Sight Imaging (NLOS) addresses the challenge of capturing a scene which is not directly visible from a sensor line-of-sight, obscured by an occluder. This work uses time-of-flight formulation to propose a learning-based method that represents implicitly the hidden scene and explores new visibility features coherent with these representations.

Introduction

Transient (also time-resolved) Imaging is a recently coined term referring to a research field focusing on capturing light at ultrafast temporal resolution. Without assuming that light speed is infinite, it is possible to capture –and simulate- light transport in transient state. The information provided by these frameworks can be exploited for NLOS Imaging by the time-of-flight contribution of photon counts in different light paths, to partially disambiguate their diffuse bounces. The advantage of the temporal resolution permits it to recover the bounce information and therefore, the distance from surface geometry occluded by the line-of-sight of a sensor. Widely conceived as imaging around corners, and even though it has not achieved transfer in-the-wild yet, it has promising applications in rescue operations at hazardous tasks, autonomous driving or medical imaging for less invasive procedures.

The occluder in NLOS scenes consists of a diffuse wall, also called relay wall, over which light paths scatter isotropically to the occluded scene when this wall is being illuminated by a laser beam. The experimental validation of time-of-flight methods for the geometry reconstruction task in NLOS Imaging has been mainly featured by jointly simulated and experimental measurements of the impulse response function of the optical system with different canonical geometry, which commonly varies in complexity regarding details, located not far from the relay wall and also floating, to avoid reconstruction artifacts by additional noise sources, like multi-path interference. Qualitative evaluation over synthetic and simulated data is still relevant in current research to benchmark reconstruction in NLOS Imaging, despite experimental reconstructions are also being conducted, addressed by femtosecond laser and sensor devices, and thereby incurring in expensive hardware. One of the leading technologies for these purposes are Single Photon Avalanche Diodes (SPADs), which permit to time-tag photon events and record time histograms of their light paths at picosecond resolution, and are a more affordable alternative for computer vision or graphics researchers which do not have access to optics laboratories.

Like other inverse problems, adapted deep learning architectures have been proposed in related work to approach NLOS in comparison with traditional techniques [2]. Assuming its well-posedness, learning-based methods are oriented to encode the solution for the NLOS forward propagation operator implicitly by the weight values of a deep neural network. These methods open future lines of work apart from the geometry reconstruction goal influenced by related areas in computer vision and graphics, such as geometry tracking or views rendering.

Method

The neural network is inspired by a Neural Radiance Field base architecture (NeRF) [1]. Adapted to this problem, the network encodes a representation of a 5D light field consisting of a cartesian space camera position and viewing direction. The output of the trained neural network is an opacity function, emulating a monochromatic luminance factor, and reflectance, which will depend solely on the viewing direction. These functions differ from traditional Neural Radiance Fields, which render volumetrically a static scene represented by an opacity function and three separate and independent RGB contributions using traditional pixel-wise ray marching.

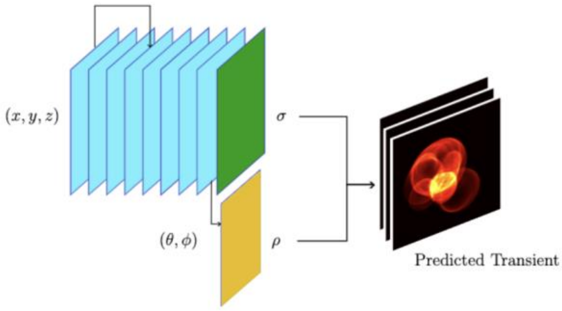

Figure 1. NeRF-like network. This deep neural network represents implicitly the hidden scene in a bounded volume, and at inference time forwards coordinates targeted to estimate the contribution of each independent voxel for the opacity function.

These two functions are jointly trained by an unbiased estimator of the transient impulse function for the optical system, compliant with the theoretical framework of confocal NLOS histograms, and it permits representing analytically the illuminated surfaces of the hidden geometry for different canonical NLOS benchmarks.

Results and Discussion

We use the synthesis of the scene by the neural network to obtain different views of the occluded scene by our proposed image formation model, which is based on an orthographic camera. This is consistent with other state-of-the-art methods and provides more detailed features on visibility for the resulting mesh reconstruction. The optimized reconstructions become finer than time-of-flight based algorithms, even though these do not require training a deep neural network for each single scene.

Conclusions and Future Work

Future work remains open as a result of the limitations of the proposed method. These identify the need of training networks with more ability to generalise over different scenes, by relaxing the objective of reconstruction by others like geometry tracking. One possibility in this direction is to use encoder-decoder networks to learn latent representations of different geometry scenes according to the SOTA canonical benchmarks,, so as to track the hidden scene instead of evaluating the fine details of each reconstruction. This interest is supplementary to other lines focusing in real-time reconstructions, for which there are results with hardware-accelerated implementations in FPGA or by adapting radiance field rendering methods. With the base network proposed, we believe that other encodings for the light field input would provide high-frequency detailed reconstructions in this context, mainly based on spatial hashing. These hashes have been validated empirically to retrieve reconstructions of high-frequency geometry efficiently by individual performance analysis of different graphics primitives.

Acknowledgements

This work has received funding from the European Union’s EUROPEAN DEFENSE FUND, Grant Agreement No. 101103242.

REFERENCES

[1] Shen, S. et al.,"Non-line-of-Sight Imaging via Neural Transient Fields". IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 43, no. 07, pp. 2257-2268, 2021. doi: 10.1109/TPAMI.2021.3076062

[2] Chen, W. et al., “Learned feature embeddings for non-line-of-sight imaging and recognition”. ACM Trans. Graph, vol. 39, no. 6, Article 230, December 2020, pp 1-18, 2020. doi:10.1145/3414685.3417825