



Multimodal Neural Radiance Field for In-Hand Robot Manipulation Tasks of Transparent Objects

Camille Taglione, Carlos Mateo, Christophe Stolz

► To cite this version:

Camille Taglione, Carlos Mateo, Christophe Stolz. Multimodal Neural Radiance Field for In-Hand Robot Manipulation Tasks of Transparent Objects. Journées Nationales de la Recherche en Robotique (JNRR 2023), Oct 2023, Moliets, France. . hal-04255537

HAL Id: hal-04255537

<https://hal.science/hal-04255537>

Submitted on 7 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

Multimodal Neural Radiance Field for In-Hand Robot Manipulation Tasks of Transparent Objects

Camille Taglione¹, Carlos M Mateo¹, Christophe Stolz¹

¹*VIBOT-ImViA Université de Bourgogne, 12 Rue de la Fonderie, 71200 Le Creusot, France*

Keywords: Neural radiance field, Multi-modal polarimetric perception, Transparent object.

In recent years, there has been significant progress in robot manipulation research, but challenges persist when it comes to tracking transparent objects using RGB-D sensors. To address this, polarimetric imaging technology has gained popularity due to cost-effective sensors becoming more accessible. This technology, which analyzes the polarization properties of light, offers valuable insights into object properties, aids in material differentiation, depth estimation, and precise pose estimation, even in challenging lighting conditions. A novel multimodal perception approach is introduced in this study, utilizing a learning-based neural radiance field method based on multimodal data to overcome the limitations of RGB-D imaging when dealing with transparent objects during in-hand manipulation (Kerr et al.).

Transparent objects present an important challenge in the field of computer vision and robotics due to their optical characteristics. Over time, considerable progress has been achieved in tackling this problem, and a promising path might be the adoption of polarimetric imaging (Dave et al.). The availability of new polarimetric sensors has paved the way for real-time polarimetric imaging. This technology leverages light’s polarization properties to provide valuable insights into the characteristics and spatial orientation of transparent objects. It has been effectively employed in tasks such as semantic segmentation of specular surfaces for autonomous driving, enhancing image quality and object pose estimation in challenging low-light conditions, reconstructing surface shapes, and even classifying object materials based on their refractive index. When capturing polarimetric images with a DoFP sensor, they initially consist of 12 channels in a raw format. To make these images usable, a preprocessing step known as demosaicing is applied, resulting in four distinct color images corresponding to different polarizer angles. These four images can then be transformed into Stokes images, providing crucial polarization information. From the Stokes image, two key pieces of information are derived: the degree of polarization (DoP) image, and the angle of polarization (AoP) image. Both of these resultant images provide valuable physical information, making polarimetric imaging a powerful tool for diverse applications.

We propose a novel approach to exploit the power of the Neural Radiance Field (NeRF) to achieve rendering in 5D, allowing us to capture and reproduce challenging transparent objects during manipulation tasks. Our method (Fig.1) uses polarimetric information such as AoP and DoP, and depth information from a depth sensor. The process begins with the extraction of depth images from the RGB source, AoP, and

DoP data. Using all the depth images, we construct a fused point cloud, effectively capturing the intricate geometry of the transparent object. Furthermore, we employ a pose estimator network to derive the object’s pose from the depth images. This pose estimation is invertible as it enables us to virtually compute the camera’s position relative to the object, and determine the viewing direction for every pixel with respect to the camera. The calculated viewing directions, along with the fused point cloud, are then fed into the Neural Radiance Field network. We employ an L2 loss for the 2D images and the Chamfer distance metric for the 3D point cloud, ensuring high fidelity in both the rendered 2D images and the 3D point cloud representation. The output of this network is a 5D rendering, promising future results that provide an accurate 3D point cloud representation as well as an accurate 2D render image of the intricate and challenging transparent object during in-hand manipulation tasks. Polarimetric

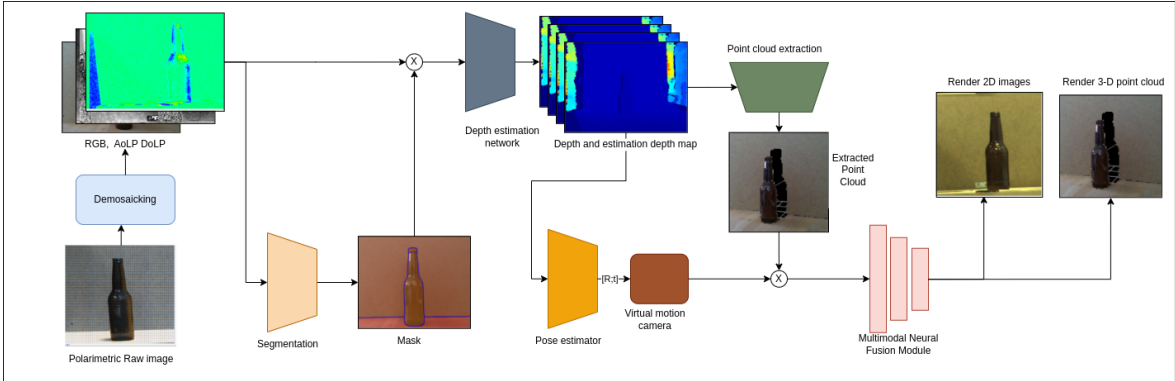


Figure 1: Framework of the method

imaging has emerged as a promising solution for transparent volumetric rendering for many different tasks. Specially, our initial results point in this direction for in-hand manipulation tasks. These improvements can be due because of the complementary nature of those two sources of information, infrared and polarize-based sensors. Our approach aims to use the polarization properties of light to overcome the challenges posed by transparent objects, giving useful information about their shape, material, and position. The use of semantic masks generated by neural networks, along with the incorporation of temporal consistency through bidirectional LSTMs, further contributes to the effectiveness of our method. In the future, the integration of polarimetric imaging into more complex tasks holds great potential for improvement in the field of autonomous robotic manipulation.

References

- Justin Kerr, Letian Fu, Huang Huang, Yahav Avigal, Matthew Tancik, Jeffrey Ich- nowski, Angjoo Kanazawa, and Ken Goldberg. *Evo-nerf: Evolving nerf for sequen- tial robot grasping of transparent objects*.
- Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. *Pandora: Polarization- aided neural decomposition of radiance*. 2022.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. *Nerf: Representing scenes as neural radiance fields for view synthesis* 2020.