ModalNeRF: Neural Modal Analysis for Free-Viewpoint Navigation in Dynamically Vibrating Scenes

(Masters or last year Engineering internship)

George Drettakis, Guillaume Cordonnier, GRAPHDECO, Inria Sophia Antipolis (France)
http://team.inria.fr/graphdeco

George.Drettakis@inria.fr
http://www-sop.inria.fr/members/George.Drettakis/
Guillaume.Cordonnier@inria.fr
http://www-sop.inria.fr/members/Guillaume.Cordonnier/

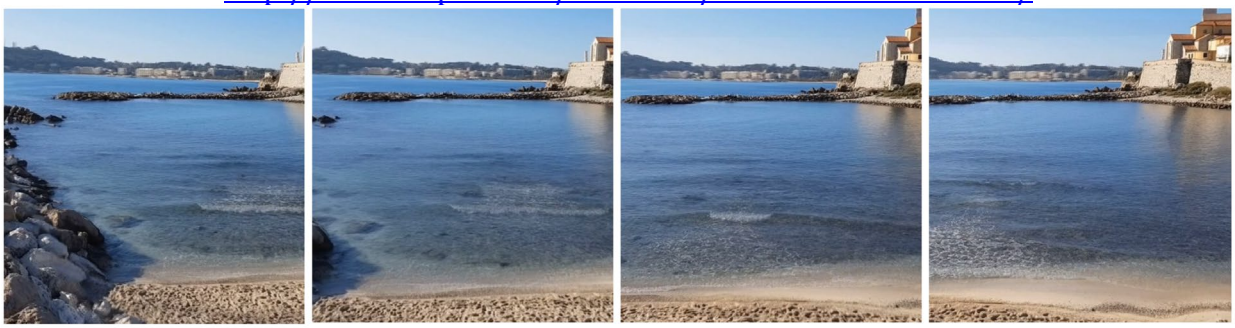

Figure 1: Snapshots of the reconstruction of a 4D scene (space + time), obtained from several decorrelated videos of the Plage de la Gravette (Antibes) [Thonat21]*. The objective of this internship is to improve the reconstruction quality thanks to neural representations, and embed physical knowledge of the motion to allow for dynamic interactions with the scene.*

Context and goal

Image Based Rendering (IBR) is an alternative to the traditional graphics pipeline where 3D scenes are reconstructed from a set of input images, and re-rendered from new viewpoints [Chaurasia13, Hedman18]. This approach allows for very efficient and extremely realistic renderings, but is traditionally limited to static scenes and viewpoints that are closed to the one of the input pictures.
In a recent approach (Figure 1) we proposed to move this paradigm to videos, turning unsynchronized input videos taken at different times from different viewpoints, into a dynamic 4D (3D + time) scene. We see two main issues in this previous work: the reconstruction quality is lower that for static scenes and it is impossible to interact with the dynamic objects.

Neural representations and rendering, and especially the *NeRF* family of algorithms [Mildenhall20], have recently revolutionized IBR by achieving unpreceded reconstruction and rendering quality. Recent solutions start to focus on dynamic scenes [Park21a,21b], albeit with very simple motion. Our goal is to extend neural representation by encoding knowledge about the physical properties of the object, which will not only allow complex animations to be captured, but will allow interactions with them based on physical forces [Davis15].

Approach

As a first experiment, we will focus on a scene with vegetation (trees or bushes) moving under a light wind. Vegetation can be captured to some extent in a controlled environment, and exhibits interesting physics. Our goal is to encode the physical dynamics of the scene in a neural representation, allowing interaction via physical forces.

We will start by encoding such scenes within a state-of-the-art dynamic neural representation [Parl21a, 21b]: these are neural networks that learn the appearance of any point in the continuous 3D domain, as well as their deformation over time, typically by maintaining a "template" and then mapping motions in other frames back to the template.

Once we have the dynamic neural representation, we use it to generate multiple (input) views for a given set of frames, and extract the physical properties of the vegetation via modal decomposition [Davis15]: we assume that Newton's second law has sinusoidal solutions and we will compute the phase and amplitude of these functions from the motion of an object in a still image. Direct editing should be possible each such view; however; achieving a multi-view consistent modal analysis will be a challenge.

In a subsequent step, we will propagate this modal representation to the dynamic neural representation, resulting in a 3D reconstructed scene that embeds physical properties of the objects and therefore enable direct interactions and editing, directly in the neural network
.
Finally, we will evaluate our approach and compare it with other strategies, e.g., constraining the possible motions by learning over a dataset of videos [Blatmann21].

Work environment and requirements

The internship will take place at Inria Sophia Antipolis in the GRAPHDECO group (http://team.inria.fr/graphdeco). Inria will provide a monthly stipend of around 1100 euros for EU citizens in their final year of masters, and ~600 euros for other candidates.

Candidates should have strong programming and mathematical skills as well as knowledge in computer graphics, geometry processing and machine learning, with experience in C++, OpenGL and GLSL on the graphics side, and tensorflow/pytorch for learning.

References

[Thonat21] T. Thonat, Y. Aksoy, M. Aittala, S. Paris, F. Durand, and G. Drettakis, "Video-Based Rendering of Dynamic Stationary Environments from Unsynchronized Inputs," *Comput. Graph. Forum (Proceedings Eurographics Symp. Render.*, vol. 40, no. 4, 2021,

[Mildenhall21] 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," 2020.

[Park21a] K. Park *et al.*, "HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields," *arXiv Prepr. arXiv2106.13228*, 2021.

[Davis15] A. Davis, J. G. Chen, and F. Durand, "Image-Space Modal Bases for Plausible Manipulation of Objects in Video," *ACM Trans. Graph.*, vol. 34, no. 6, 2015,

[Park21b] K. Park *et al.*, "Nerfies: Deformable Neural Radiance Fields," *ICCV*, 2021.

[Blattmann21] A. Blattmann, T. Milbich, M. Dorkenwald, and B. Ommer, "iPOKE: Poking a Still Image for Controlled Stochastic Video Synthesis." 2021.

[Chaurasia13] G. CHAURASIA, S. DUCHENE, O. SORKINE-HORNUNG, & G. DRETTAKIS. (2013). Depth synthesis and local warps for plausible image-based navigation. *ACM Transactions on Graphics (TOG)*, *32*(3), 30. http://www-sop.inria.fr/reves/Basilic/2013/CDSD13/

[Hedman18] P. HEDMAN, T. RITSCHEL, G. DRETTAKIS, G. BROSTOW, Scalable Inside-Out Image-Based Rendering, ACM Transactions on Graphics 35 (SIGGRAPH Asia), 6, December 2016 http://www-sop.inria.fr/reves/Basilic/2016/HRDB16/