Generalization of X-fields image interpolation model to higher dimensions

Jack Wang — Groton School
Mentor: Guillaume Leclerc

Overview

- Motivation
- Related Work
- My work
- Findings
- Future plans

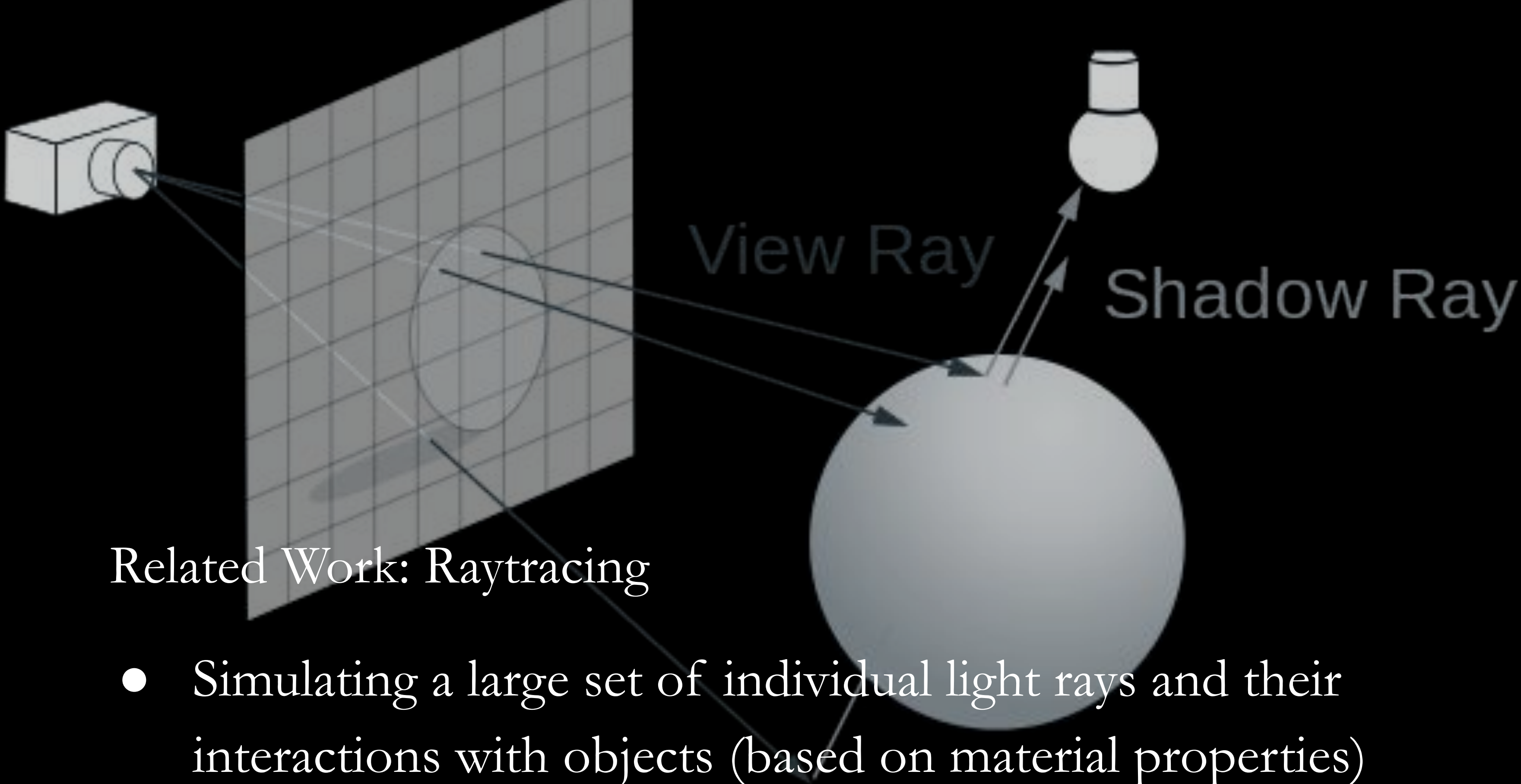
Where are realistic renders used?

- Video games

Films (especially animated ones)

Creation of debug models

- Finding the worst case scenarios for a machine learning algorithm, with the intent of fixing the performance



Related Work: Raytracing

- Simulating a large set of individual light rays and their interactions with objects (based on material properties) from an imaginary "eye"

Unstructured Lumigraph Rendering (ULR): create proxy geometry to warp multiple images into a target view and blend them with corresponding weights

Volumetric occupancy representation: avoids explicit depth reconstruction, which allows for softer and better results

Related Work: X-Fields

- "X-Field" (where X may be any combination of view, light, time, or other dimensions)
- NN representation that generalizes entire geometry, motion, and illumination changes
- Implicit map such that for any view, light, or coordinate, it can quantify how it will move if view, time or light coordinates change for any pixel
- Fully differentiable
- Independent interpolation of appearance, albedo, and shading (Hadamard product)

$$\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^{n_\mathrm{d}}$$

$$L_{\mathrm{out}}(\mathbf{x}) = \mathtt{int}(A(L_{\mathrm{in}}(\mathbf{y})), \mathbf{y} \to \mathbf{x}) \odot \mathtt{int}(E(L_{\mathrm{in}}(\mathbf{y})), \mathbf{y} \to \mathbf{x})$$

$$L_{\mathrm{out}}^{(\theta)}(\mathbf{x}) \in \mathcal{X} \to \mathbb{R}^{3 \times n_\mathrm{p}}$$

My work

- Changed the coordinate input method to make it more convenient and usable
- Added another coordinate Z to construct a 3D view interpolation instead of the 2D one from the original model
- Generalized for most real-world use cases
- Added ability to fix a coordinate in generation, allowing for more specific interpolation videos
- Generation of 3DB (program designed to find failure modes) tugboat datasets
- Research into the tradeoffs

512p input, 512p output, 15x15x15 image dataset

2160p input, 560p output, 15x15x15 image dataset

240p inputs and outputs, 10x10x10 dataset

Findings

- The revised X-Fields can generate promising interpolation results with relatively sparse datasets and with large view angle changes.
- Parameters such as learning rate and the bandwidth parameter in soft blending have impacts over the interpolation quality and construct trade-offs between training cost and interpolation quality.
- Certain backgrounds added to reference images can pose a challenge for interpolation.

Future plans

- Combining Transformer Networks (using the idea of attention, differentially weighting the importance of each part of the input data) and X-Fields' hard-coded graphics
- Generation and testing of more datasets and objects

New X-Fields

CAIN

Thanks to...

My mentor, Guillaume Leclerc

Dr. Slava Gerovitch and Professor Devadas

MIT PRIMES + CSAIL

Thanks for listening!

Questions?