4K4D Lightweight Offline Deployment on VR Devices

Research Report - 1

Zuyu Guo

March 18, 2025

Northeastern University

Table of contents

Background

Research Background

Real-Time 4D View Synthesis at **4K Resolution**

- What: Generate high-resolution 3D scenes of arbitrary viewpoints from multi-view videos.
- Why: The speed of the traditional method is still limited when rendering high-resolution images.
- How: hardware-accelerated rasterization + Pre-estimate colors

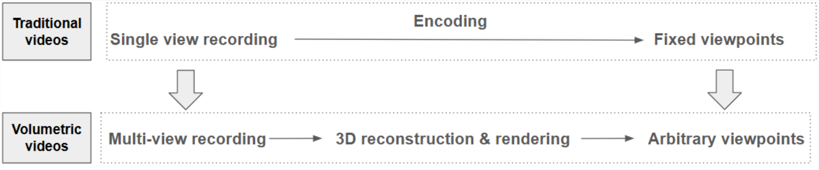

Figure 1: Traditional Video and Volumetric Video

Related Work

Related Work: NeRF

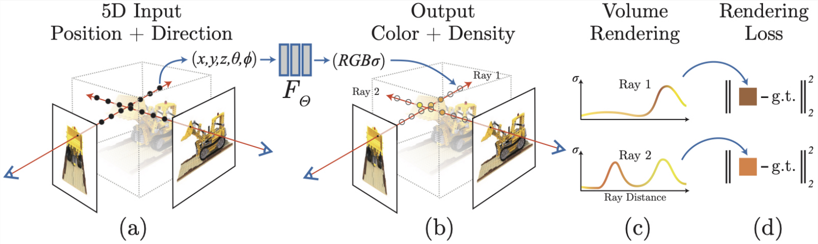

Figure 2: NeRF: Representing Scenes as Neural Radiance Fields

NeRF learns an implicit 3D scene representation by MLP to map spatial coordinates and viewing directions to volume density and radiance.

$$F_\theta : (x, y, z, \theta, \phi) \xrightarrow{\text{MLP}} (\sigma, c) \tag{1}$$

where (x, y, z) represents spatial coordinates, (θ, ϕ) is the viewing direction, σ is the volume density, and c is the color.

Related Work: DyNeRF

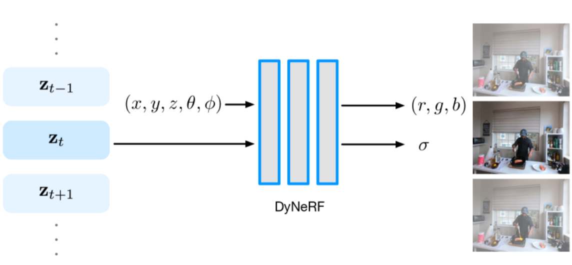

Figure 3: DyNeRF: Dynamic Neural Radiance Fields

DyNeRF extends NeRF by introducing a **temporal dimension** t to model dynamic scenes. It learns an implicit 4D representation by using an MLP.

$$F_\theta : (x, y, z, t, \theta, \phi) \xrightarrow{\text{MLP}} (\sigma, c) \tag{2}$$

where time t is introduced to allow the neural network to learn time-dependent changes in volume density σ and radiance c.

Related Work: K-Planes

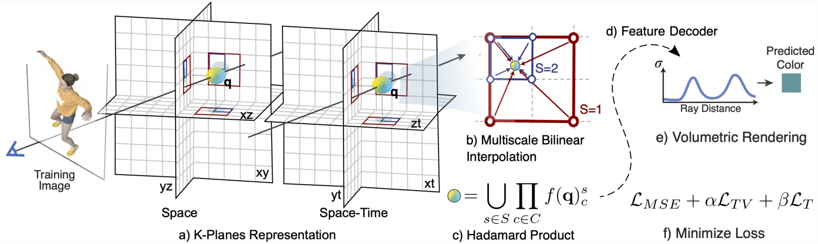

Figure 4: K-Planes: Explicit Radiance Fields in Space, Time, and Appearance

K-Planes introduces an explicit 4D radiance field representation by factorizing the scene into multiple 2D feature planes along different axes.

$$f(x, y, z, t) = \theta_{xy}(x, y) \oplus \theta_{xz}(x, z) \oplus \theta_{yz}(y, z) \oplus \theta_{tx}(t, x) \oplus \theta_{ty}(t, y) \oplus \theta_{tz}(t, z) \tag{3}$$

$$(\sigma, c) = \mathrm{MLP}(f(x, y, z, t)) \tag{4}$$

K-Planes directly queries precomputed feature grids and combines them through a small MLP to produce volume density and color.

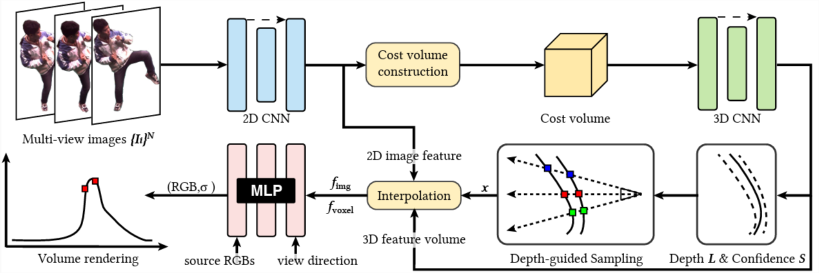

Figure 5: Efficient Neural Radiance Fields

ENeRF improves the efficiency of Neural Radiance Fields by introducing sparse ray sampling and a more compact MLP architecture.

$$F_\theta : (x, y, z, \theta, \phi) \xrightarrow{\text{Sparse Sampling + MLP}} (\sigma, c) \tag{5}$$

Optimization Analysis

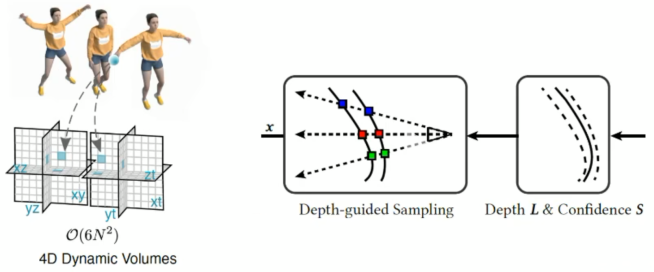

Figure 6: Optimization Route

Two main approaches to speed up:

1. **Reducing inference cost: K-Planes**
   - The feature planes store explicit information, allowing fast lookup and reducing the number of MLP evaluations.
2. **Reducing sampling points: ENeRF**
   - ENeRF uses an adaptive sampling strategy to focus on important regions, reducing the number of MLP evaluations.

4K4D Approach

4K4D Core Ideas

Core Idea: Reduce Inference Cost & Accelerate Sampling Process

- **Rendering Cost:**

$$\text{Rendering Cost} = \text{Number of samples} \times \text{Network Inference Cost} \quad (6)$$

- **Two Key Optimizations:**
  1. **Hardware-accelerated Sampling**
     - Use GPU-based rasterization and differentiable depth peeling to reduce redundant ray sampling.
  2. **Precomputed Representations for Real-time 4K Rendering**
     - Store explicit radiance field properties, including precomputed color representations, to avoid redundant computations.

4K4D pipeline

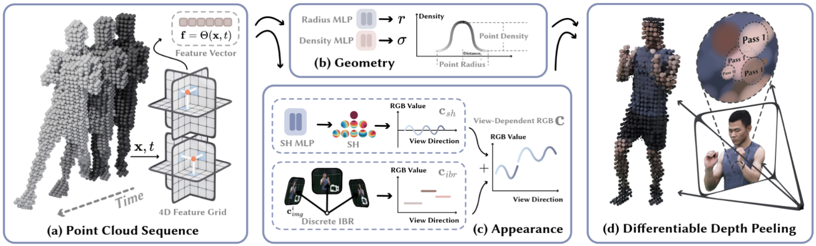

Figure 7: 4K4D Model

4K4D pipeline

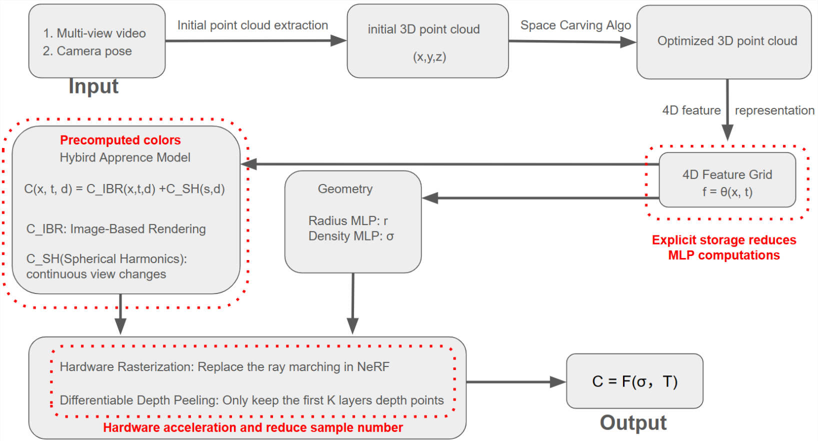

Figure 8: 4K4D Pipeline

Experiment

Experiment

Current Progress

- Successfully run the ENeRFi(ENeRF Improved) method within the EasyVolcap framework.

Current Problems:

- Large memory GPU is Required. (24 GB RTX 4090)
- EasyVolcap Framework must be run with the GUI, only the server is not enough.
- 4K4D real-time rendering is not end-to-end real-time.
- Storage costs are too high.

Question

Some thoughts on the project

- Pre-training work is extremely complex and time-consuming and may take several days for a normal dataset.
    - Cloud Computing + Distributed Data Storage + high speed transfer.
    - Only perform final rendering on VR devices
- Model Optimization: Knowing the data depth information can simplify the model structure in terms of calculation and storage
- System Development:...

Questions?

References i

[1] [2] [3] [4] [5]

S. Fridovich-Keil, G. Meanti, F. Warburg, B. Recht, and A. Kanazawa.
K-planes: Explicit radiance fields in space, time, and appearance, 2023.

T. Li, M. Slavcheva, M. Zollhoefer, S. Green, C. Lassner, C. Kim, T. Schmidt, S. Lovegrove, M. Goesele, R. Newcombe, and Z. Lv.
Neural 3d video synthesis from multi-view video, 2022.

H. Lin, S. Peng, Z. Xu, Y. Yan, Q. Shuai, H. Bao, and X. Zhou.
Efficient neural radiance fields for interactive free-viewpoint video, 2022.

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.

Z. Xu, S. Peng, H. Lin, G. He, J. Sun, Y. Shen, H. Bao, and X. Zhou.
4k4d: Real-time 4d view synthesis at 4k resolution, 2023.