A High-Performance Sparse Voxel Grid Framework for Intrinsic 3D Reconstruction on GPUs

Zihang Zhang* and Fumihiko Ino

Graduate School of Information Science and Technology, The University of Osaka

*email: zh-zhang@ist.osaka-u.ac.jp

Introduction

Background

- **Intrinsic decomposition** is a fundamental task in inverse rendering, aiming at **estimating scene attributes** from observed images.

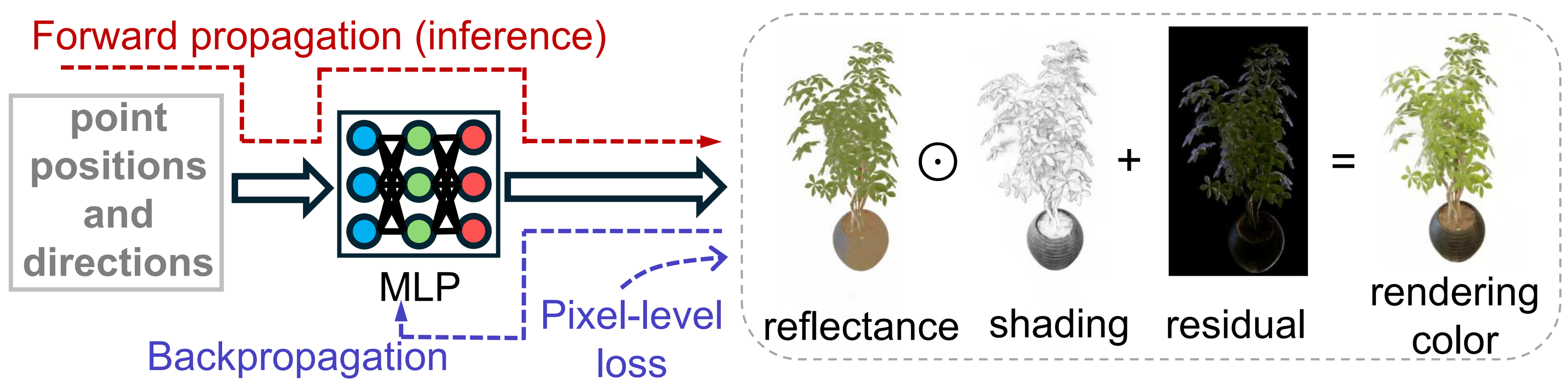
$$\text{RGB color} = \text{reflectance} \odot \text{shading} + \text{residual} \quad (1)$$

- This decomposition enables **independent editing** of material and lighting.
- **reflectance: base color; shading: illumination; residual: specular lighting.**

Motivations & Goals

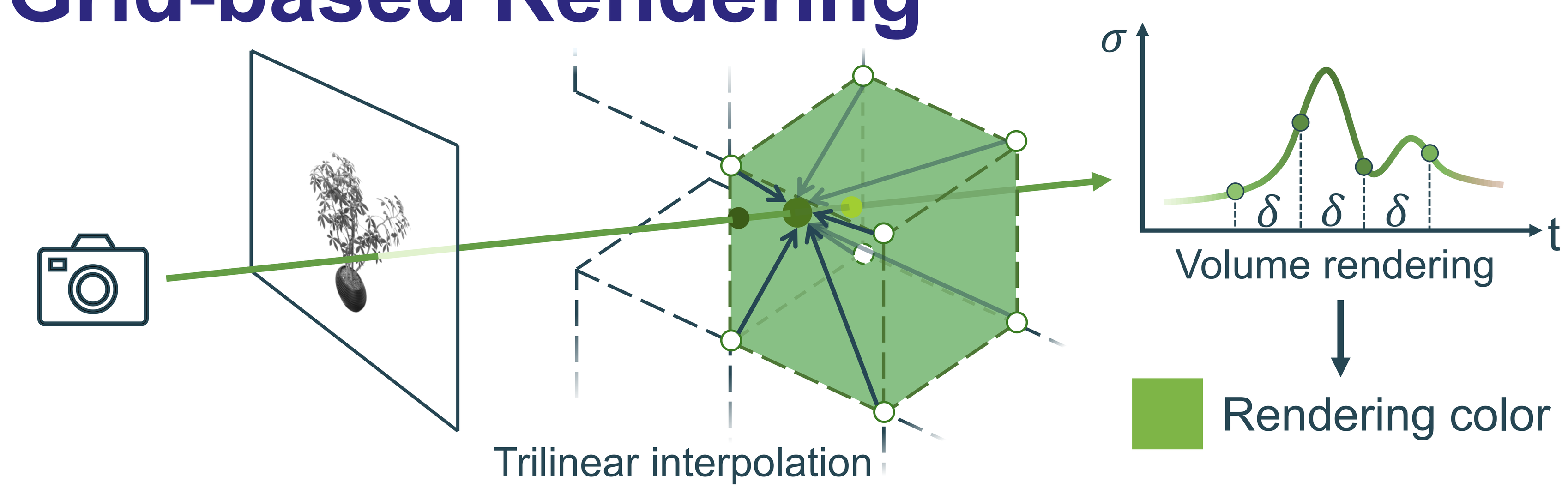
- Existing neural methods [1] represent scene attributes via neural networks (MLPs), resulting in **slow inference speeds** and **limited parallelism**.
- We propose IntrinsicGrid to replace MLPs with **Sparse Voxel Grids**, enabling direct memory access and mapping rays to GPU threads for **significant acceleration**.

Neural Rendering & Decomposition



- Neural network-based methods [1] use MLPs to <u>implicitly</u> **store and represent** scene attributes.
  - **High latency:** A **complete MLP forward propagation** is required to sample a single point due to **sequential layer dependencies**.
  - **Lack of spatial information:** Operate in 2D image space or latent spaces, **losing spatial information** necessary for decomposition.

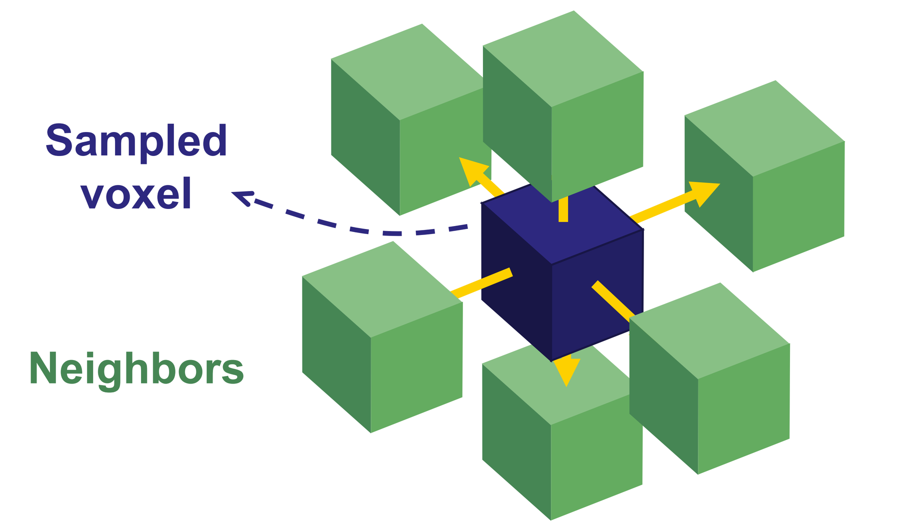
Grid-based Rendering



- Grid-based methods (inspired by Plenoxels [2]) <u>explicitly</u> store scene attributes in 3D voxel grids.
  - **Fast sampling: Directly access grid voxels** to get stored sampling point values without neural network overhead.
  - **Spatial coherence:** Enable **efficient** spatial constraints on **spatially neighboring voxels** via grid indexing.

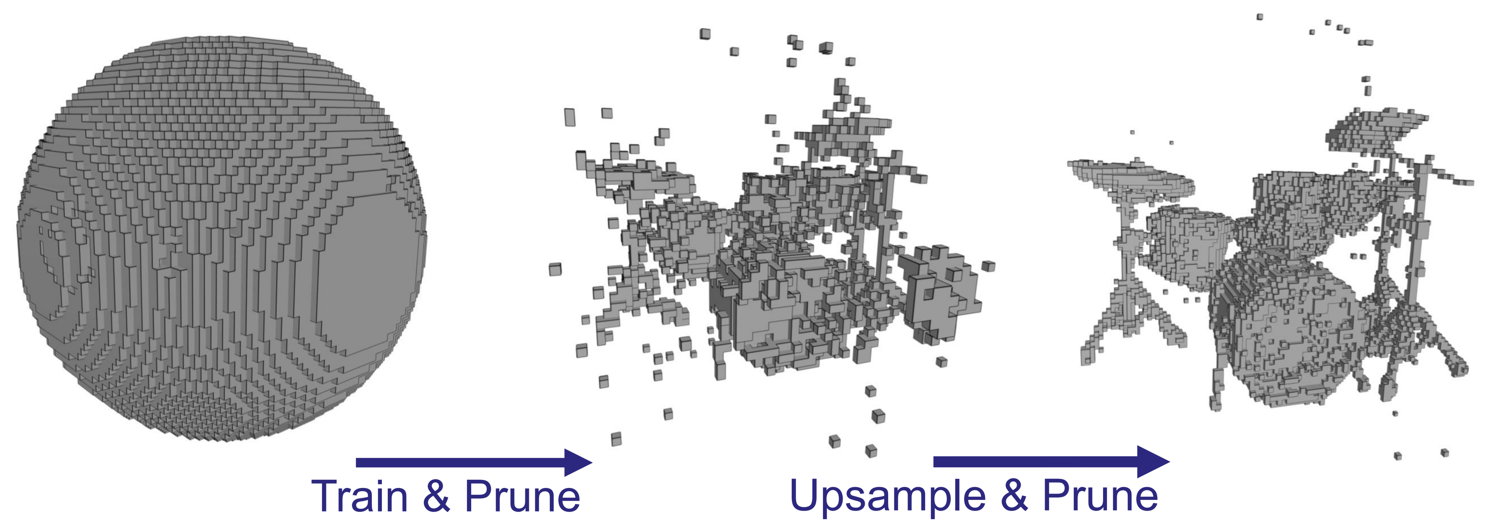
Proposed Framework: IntrinsicGrid

1. Spatial Constraints



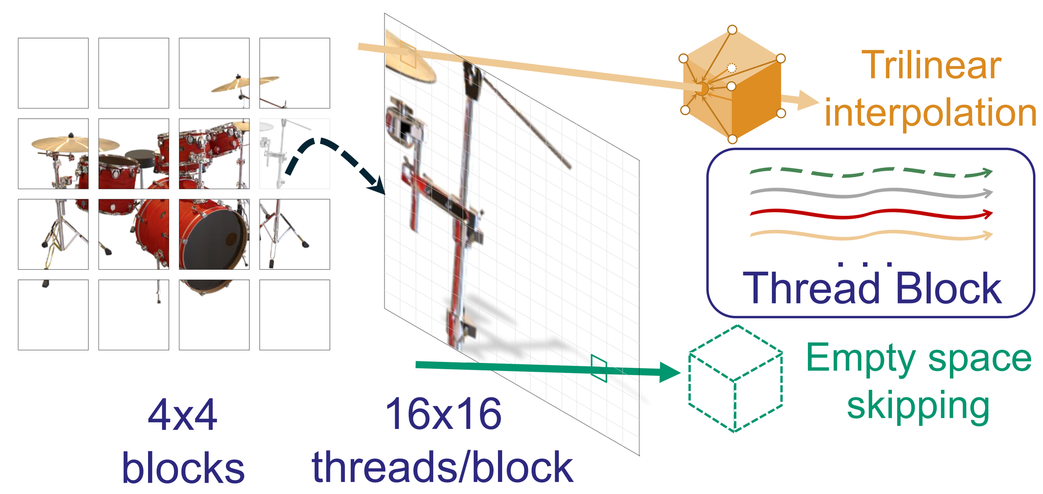
- **Randomly sample 1% of the voxels.**
- Apply spatial smoothness constraints to **the voxels** and **their 6 neighboring voxels** to address decomposition ambiguity within acceptable overhead.

2. Coarse-to-Fine Sparse Grid Optimization



- We **optimize the target scene attributes** (reflectance, shading, and residual) **stored in sparse voxel grid** without any neural networks.
- **Training starts at a low resolution** to reduce computation.
- **Prune empty voxels** to maintain grid sparsity.

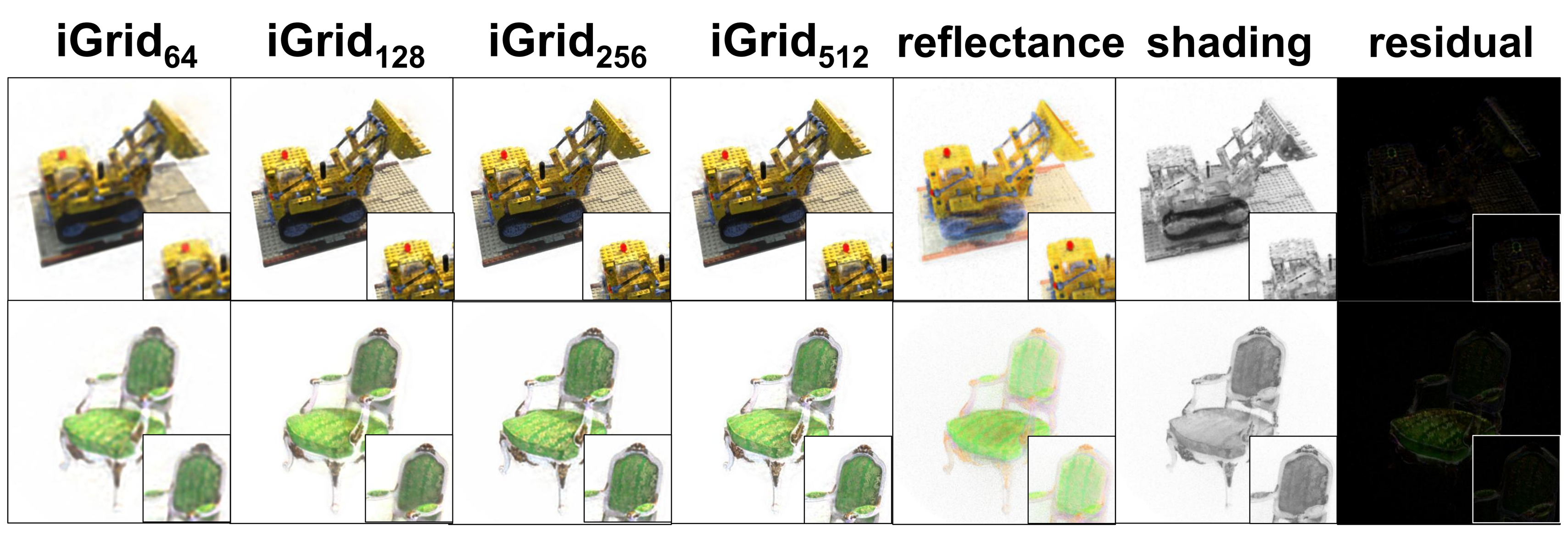
3. Custom CUDA Acceleration



- **Each sampling ray** is assigned to one GPU thread for **independent processing**.
- Ray marching is **fully on GPU**, eliminating data transfer overhead.
- <u>**Only for forward propagation.**</u>

Experimental Results

Metrics	iNeRF	$iGrid_{64}$	$iGrid_{128}$	$iGrid_{256}$	$iGrid_{512}$	SVOX2
Training (h)	7.25	0.35	1.17	2.41	4.37	0.12
Rendering (s)	6.83	0.23	0.41	0.66	1.45	—
Speedup	1×	29.7×	16.7×	10.3×	4.7×	—
CUDA Render (ms)	—	2.75	6.06	19.82	49.10	43.83
CUDA Speedup	—	2484×	1127×	345×	139×	—
PSNR↑ (avg.)	29.94	24.36	26.96	28.07	28.46	30.98



- **Setup**: Evaluated on the *NeRF-Synthetic* dataset using a single NVIDIA RTX 6000 Ada with PyTorch 2.9.0, CUDA 12.8, and CUDA toolkit 12.8.
- **IntrinsicNeRF: iNeRF [1].**
- **IntrinsicGrid: iGrid (ours).**
- **SVOX2: Plenoxels [2]** with resolution 512, a sparse voxel grid-based rendering method.

- iGrid <u>renders over two orders of magnitude faster</u> than iNeRF.
- iGrid has <u>comparable rendering quality</u> to iNeRF.
- Choose suitable grid resolution for downstream tasks based on the requirements and hardware capability.

Conclusion

- The results indicate that intrinsic decomposition-based 3D reconstruction can be accelerated using sparse voxel grids to exploit the <u>massive parallelism</u> **of modern GPUs**.
- IntrinsicGrid achieves over <u>139× faster rendering</u> than IntrinsicNeRF with <u>moderate quality trade-off</u> (PSNR 28.46 vs 29.94), making it suitable for interactive editing applications.
- These findings highlight the potential of **IntrinsicGrid as a <u>general and efficient framework</u> for evaluating diverse decomposition assumptions** and accelerating inverse rendering pipelines.

Acknowledgements

This study was in part supported by the Japan Society for the Promotion of Science KAKENHI grant number 23K28061.

References

[1] Weicai Ye, Shuo Chen, Chong Bao, Hujun Bao, Marc Pollefeys, Zhaopeng Cui, and Guofeng Zhang. 2023. IntrinsicNeRF: Learning Intrinsic Neural Radiance Fields for Editable Novel View Synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 339–351

[2] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Net-works. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5501–5510.

[3] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Proceedings of the European Conference on Computer Vision. 405–421.