

Dynamic Scene Modeling and Rendering: A Survey of Methods and Applications

Yuwei Zhao^{1,2}, Kaiyuan Zhang^{1,2}, Yuxiang Liu^{1,2}, Yilin Zhang^{1,2}, and Keqin Zhang^{1,2}

Abstract—Dynamic scene modeling and reconstruction from video streams have witnessed a paradigm shift with the advent of Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS). While NeRF-based methods offer photorealistic quality via implicit neural representations, they often suffer from prohibitive training and rendering costs. Recently, 3DGS has emerged as a powerful alternative, enabling real-time rendering through explicit splatting techniques, though it introduces new challenges regarding storage overhead and temporal consistency. In this paper, we present a comprehensive survey of video-based 3D reconstruction methods, categorizing them into implicit and explicit approaches. We systematically evaluate their trade-offs in terms of rendering quality, speed, and memory consumption across diverse scenarios. Furthermore, we summarize the limitations of current state-of-the-art methods and discuss future directions. Our compiled resources and code are available at <https://github.com/8barbatos/Academic-English-Group-Paper>.

Index Terms—3D Reconstruction, Dynamic Scene Modeling, Neural Radiance Fields (NeRF), Gaussian Splatting, Novel View Synthesis, Survey.

I. INTRODUCTION

WITH the rapid advancement of virtual reality (VR), augmented reality (AR), and the Metaverse, the demand for photorealistic 3D content creation has surged exponentially. Dynamic scene modeling, which aims to reconstruct 3D geometry and appearance from 2D video streams, serves as a fundamental technology for these applications, enabling immersive telepresence, free-viewpoint video, and digital human avatars [1]. Unlike static scene reconstruction, modeling dynamic scenes from video presents a highly ill-posed inverse problem due to the entanglement of object motion, topology changes, and time-variant lighting conditions.

Traditionally, 3D reconstruction relied on Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithms [2]. While these methods, such as COLMAP, provide robust camera pose estimation, they struggle to capture thin structures and view-dependent effects (e.g., reflections). More importantly, traditional pipelines typically assume a static world, making them brittle when applied to dynamic video sequences where geometry consistency is violated over time.

The field witnessed a paradigm shift with the introduction of **Neural Radiance Fields (NeRF)** [3]. By representing scenes as implicit continuous functions parameterized by Multi-Layer Perceptrons (MLPs), NeRF achieved unprecedented rendering quality. Subsequent works, such as D-NeRF [4] and Nerfies [5], extended this implicit paradigm to dynamic domains by

introducing deformation fields to handle non-rigid motion. However, implicit methods suffer from prohibitive computational costs due to the extensive ray-marching sampling required during both training and inference, limiting their deployment in real-time applications.

Recently, **3D Gaussian Splatting (3DGS)** [6] has emerged as a compelling alternative, marking a return to explicit volumetric representations. By combining the differentiability of deep learning with the efficiency of rasterization-based rendering, 3DGS enables real-time rendering (100+ FPS) and fast training speeds. This breakthrough has triggered a new wave of research focused on extending Gaussian primitives to 4D spatiotemporal modeling [7], [8], aiming to combine the efficiency of explicit representations with the flexibility of neural fields.

Despite the explosion of research papers in this domain, a systematic comparison between implicit (NeRF-based) and explicit (Gaussian-based) approaches for video modeling is lacking. In this survey, we provide a comprehensive review of the state-of-the-art methods for video-based 3D reconstruction. The main contributions of this paper are summarized as follows:

- We propose a structured taxonomy of dynamic scene modeling methods, categorizing them into *Deformation-based*, *Spacetime-based*, and *Hybrid* approaches across both implicit and explicit representations.
- We provide an in-depth analysis of the transition from NeRF to 3D Gaussian Splatting, highlighting the trade-offs between rendering quality, training efficiency, and storage overhead.
- We conduct a comparative evaluation of representative frameworks and discuss open challenges, including storage optimization and long-duration video modeling, to guide future research directions.

II. PRELIMINARIES

A. Mathematical Formulation

...

B. Fundamentals of Neural Radiance Fields

...

C. Fundamentals of 3D Gaussian Splatting

...

¹School of Computer Science and Technology, Ocean University of China, Qingdao 266100, China.

²School of MPs & EPs, Heriot-Watt University, Edinburgh EH14 4AS, UK.

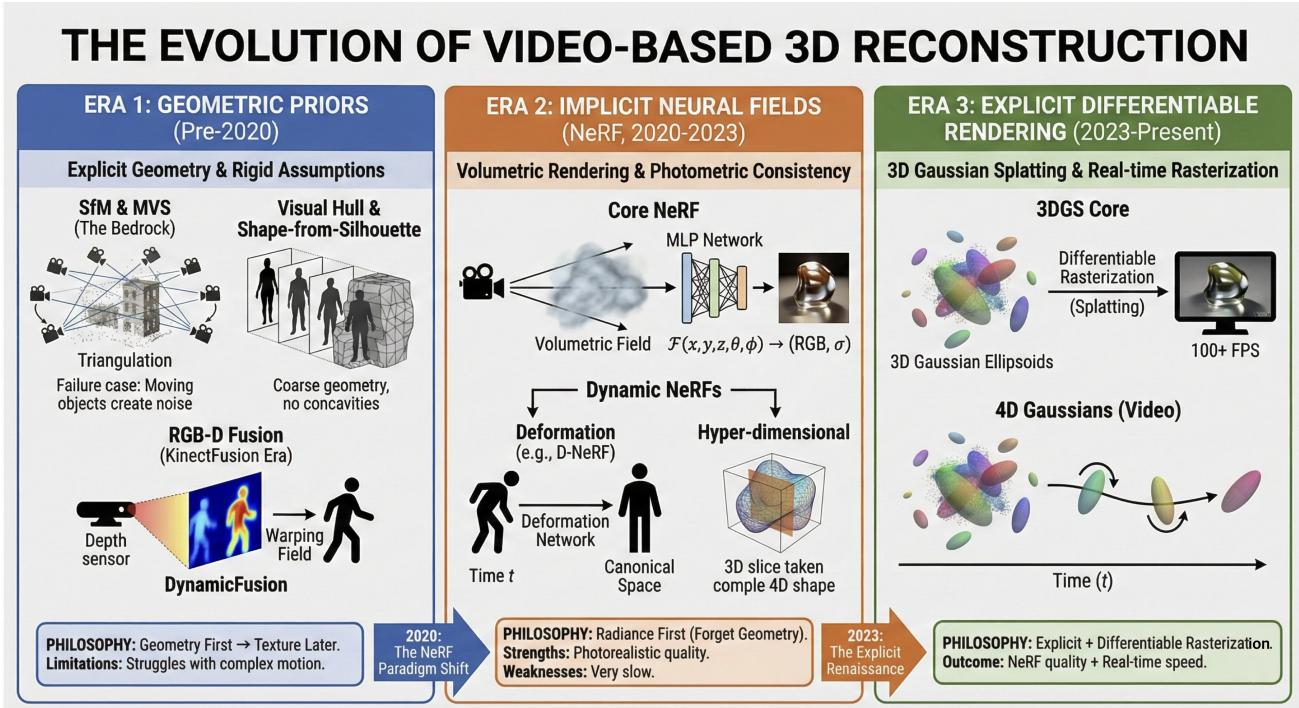


Fig. 1: Visualization of the evolution of video-based 3D reconstruction

III. TAXONOMY OF DYNAMIC SCENE MODELING

A. Implicit Deformation Fields (e.g., D-NeRF)

...

B. Spacetime Neural Fields (e.g., DyNeRF)

...

C. Dynamic Gaussian Splatting (e.g., 4D-GS)

...

D. Hybrid Representations

...

IV. DATASETS AND METRICS

A. Common Datasets (N3V, D-NeRF, etc.)

...

B. Evaluation Metrics (PSNR, SSIM, LPIPS, FPS)

...

V. COMPARATIVE ANALYSIS

A. Quantitative Benchmarks

...

B. Qualitative Visual Comparisons

...

C. Efficiency Analysis (Time & Memory)

...

VI. OPEN CHALLENGES AND FUTURE DIRECTIONS

A. Storage Efficiency and Compression

...

B. Handling Long-Duration Videos

...

C. Integration with Generative Models

...

VII. CONCLUSION

...

REFERENCES

- [1] Apple Inc., *Expression Estimation for Headsets Using Low-Profile Antenna and Impedance Characteristic Sensing*.
- [2] J. Schönberger, J.-M. Frahm, M. Pollefeys, P.-E. Sarlin, and S. Liu, *COLMAP*, [Online]. Available: <https://colmap.github.io/index.html>
- [3] 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,” *arXiv preprint*, arXiv:2003.08934, 2020. [Online]. Available: <https://arxiv.org/abs/2003.08934>
- [4] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, “D-NeRF: Neural Radiance Fields for Dynamic Scenes,” *arXiv preprint*, arXiv:2011.13961, 2020. [Online]. Available: <https://arxiv.org/abs/2011.13961>
- [5] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, “Nerfies: Deformable Neural Radiance Fields,” *arXiv preprint*, arXiv:2011.12948, 2021. [Online]. Available: <https://arxiv.org/abs/2011.12948>

- [6] J. Kerbl, A. Wang, F. Rousselle, and V. Koltun, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” *arXiv preprint*, arXiv:2304.08914, 2023. [Online]. Available: <https://arxiv.org/abs/2308.04079>
- [7] G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian, and X. Wang, “4D Gaussian Splatting for Real-Time Dynamic Scene Rendering,” *arXiv preprint*, arXiv:2310.08528, 2024. [Online]. Available: <https://arxiv.org/abs/2310.08528>
- [8] J. Luiten, G. Kopanas, B. Leibe, and D. Ramanan, “Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis,” *arXiv preprint*, arXiv:2308.09713, 2023. [Online]. Available: <https://arxiv.org/abs/2308.09713>

VIII. BIOGRAPHY SECTION

Yuwei ZHAO Use `\begin{IEEEbiographynophoto}` and the author name as the argument followed by the biography text.