Integrating Motion-Aware and Direction-Aware Techniques for Improved 3D Gaussian Splatting in Dynamic Scenes

Mai Bui

Abstract— **Accurately reconstructing 3D environments, especially with moving objects, remains a significant challenge. While advanced techniques like neural rendering (NeRF) produce detailed 3D models, they are computationally intensive, limiting real-time applications. 3D Gaussian Splatting (3D GS) offers a more efficient alternative, but struggles with motion blur in dynamic scenes. This work presents an integrated approach that combines motion-aware and direction-aware methods to enhance 3D GS. The motion-aware component uses optical flow to track object movements, while the direction-aware DaRePlane technique captures complex geometries and reduce motion blur. These are integrated through a hierarchical pipeline - first performing a coarse motion-aware reconstruction, then refining it using direction-aware processing. Adaptive splat representations and parallel processing optimize computational and storage requirements, ensuring real-time performance. Experiments show our integrated solution outperforms standalone techniques, reducing motion blur and preserving geometric fidelity, advancing the state-of-the-art in practical 3D scene reconstruction.**

Index Terms— **Gaussian Splatting ; motion-aware; direction-aware, optical flow, DaRePlane**

I. Introduction

While dynamic scene reconstruction plays an important role in medical training, autonomous driving, sport and performance exhibition, etc, there has been multiple challenges including real-time performance, scene complexity, data requirement and temporal consistency. In this paper, I want to focus on solving the problem of temporal consistency, which means the consistency of data generated at different times.

II. Existing methods & current limitations

A. Neural Radiance Fields (NeRF)

Introduced in 2020, this neural network-based approach was a breakthrough and a great benchmark for synthesizing novel views of complex scenes. It trained multiple images from different viewpoints, going through ray marching and traditional volumn rendering technique to predice volumn density and RGB color for each pixel to create 3D reconstruction of the scene [1]. Since then, there has been significant efforts to improve, producing multiple variants for different purposes: D-NeRF for dynamic scene reconstruction [2], Instant RGB for realtime rendering [3], Nerflies for handling deformable scene [4]. However, despite notable development, its time and power consuming presents great challenges. [1]. Additionally, as NeRF is frame-by-frame based training and is lack of ”recall” previous frame mechanism, this state-of-the-art approach produces noticeable artifacts in dynamic scenes, resulting in bad temporal inconsistency.

Mai Bui with Mount Holyoke College, Department of Computer Sciece, 50 College Street, South Hadley, MA USA {bui23m}@mtholyoke.edu

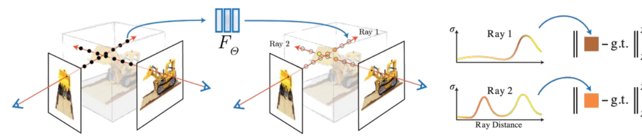

Fig. 1. Neural Radiance Fields (NeRF) Architecture and View Synthesis Workflow [1]

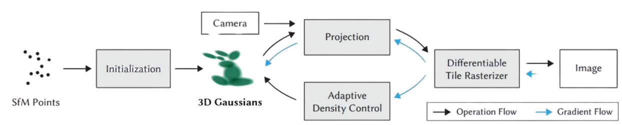

Fig. 2. Neural Radiance Fields (NeRF) Architecture and View Synthesis Workflow [5]

B. Gaussian Splatting (GS)

In order to offer realtime rendering and training which provides tremendously more practical use, 3D GS was introduced as an alternative way for 3D reconstruction. In stead of representing 3D scenes with neural fields, it uses a collection of anisotropic 3D Gaussian functions (splats). Those include information: position, covariance, color, opacity. By projecting 3D images into 2D planes, then went through point-based volumn rendering, this approach produced photorealistic images from any viewpoint [5]. However, due to its reliance on splats, in fast-moving environment, splats may shift or distort, resulting in blurry or jittery reconstruction.

C. Motion-aware

In order to solve the loss of motion information in fast paced environment, optical flow, in motion-aware methods was introduced to track object movement across frames, aiding in reducing temporal inconsistencies. However, conventional optical flow methods struggle with covariance shifts and lack direction sensitivity, limiting their effectiveness in capturing complex motions. [6]

D. Direction-aware

To overcome problem with direction sensitivity, DaRePlane introduced a direction-aware representation that captures 6 instead of 3 traditionally directions. This helps the drop from 4D to 2D less steep, lossing less information, capturing detailed information in dynammic settings and producing superior performance in novel view synthesis. [7]

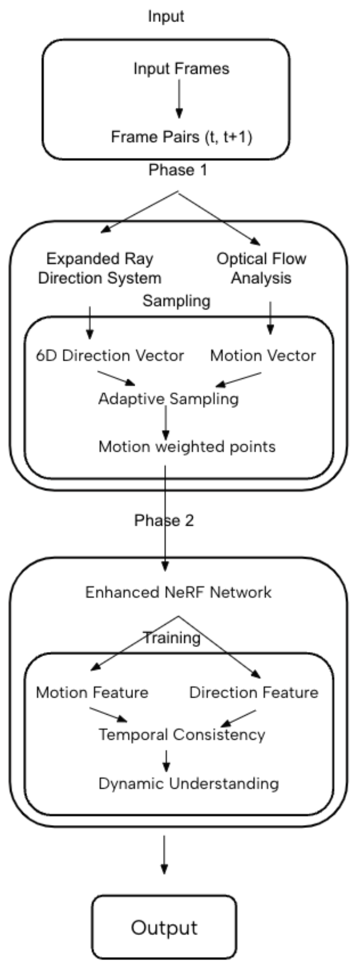

Fig. 3. Implementation plan coding workflow

III. Research Proposal

A. Integration Idea

I am interested in integration direction-aware and motion-aware methods and put it into Gaussian Splatting to evaluate the model's improvement easier and more holistic. I believe the benefit of getting detailed information from direction-aware will be a great addition/component to overcome the motion blur in motion-aware. And given motion information will contribute to provide better reasoning for transparancy and occultation of motion-aware approach. In the future, I would love to present a pipeline to ensure smooth integration between 2 approaches and motion, direction-aware with 3D GS and NeRF. In this proposal, as Google Colab is an environment with limit of GPU usage and not suitable for 3D GS's complex dependency, I focused on creating pipeline and planning for integration in NeRF.

B. Code explanation

Code can be found in this Github Link

1) Phase 1: Direction and Motion Enhancement: The first phase will focus on expanding the ray direction system from three to six dimensions, integrating up/down viewing angles and forward/backward/left/right motion predictions alongside existing spatial coordinates. This foundational change will require careful modification of the get_rays function while maintaining compatibility with existing rendering pipelines. Once the expanded directional system is stable, we'll implement motion tracking capabilities through optical flow analysis between consecutive frames - this represents a critical bridge between static and dynamic scene understanding.

More specifically, we'll create a preprocessing stage that analyzes pairs of consecutive frames to compute motion vectors - these vectors will track how each point moves from one frame to the next. A new preprocessing stage analyzes consecutive frame pairs to compute motion vectors, tracking point movement between frames. This requires the implementation of a compute_optical_flow function, which can utilize either traditional computer vision techniques or a specialized neural network for pixel-level movement estimation. Real-time performance remains a critical consideration throughout this implementation.

Building upon this foundation, we introduce an adaptive sampling strategy that revolutionizes how the system processes dynamic scenes. The render_flat_rays function undergoes enhancement to adjust sampling density based on detected motion. For instance, when capturing a scene with a moving person, the system intelligently allocates more computational resources to areas with significant movement while maintaining efficient processing of static backgrounds. This adaptive approach relies on a sophisticated weighting algorithm that considers both motion magnitude and direction.

2) Neural Network Enhancement: The second phase focuses on architectural modifications to the neural network itself. The network requires significant redesign to accommodate six directional dimensions and specialized layers for motion flow features. This enhancement enables new predictive capabilities, including velocity and acceleration estimation, along with uncertainty quantification for moving objects within scenes.

The training pipeline receives corresponding updates to support these new capabilities. Temporal consistency checks and motion-aware loss functions ensure the system learns to accurately predict and represent dynamic scene elements. To manage the increased computational complexity, we implement progressive training strategies that carefully balance performance requirements with available resources.

C. Experiments

With the theoretical foundation established, our next steps focus on practical implementation and evaluation. The immediate priority is to implement the enhanced NeRF system, starting with the expanded direction-aware components followed by motion tracking integration. I plan to conduct comprehensive experiments comparing our enhanced model against baseline NeRF implementations, particularly focusing on scenes with significant dynamic elements and complex viewing angles and using Plenoptic Video Dataset [8] and Cochlear Implant Surgery [9].

TABLE I

OVERVIEW OF EXPERIMENTAL DESIGN AND METRICS

Aspect	Details	Evaluation Metric
Dataset	Plenoptic Video Dataset, Cochlear Implant Surgery Dataset	Robustness in dynamic and medical scenes
Baseline	Vanilla NeRF implementations	Motion handling, view rendering
Enhancements	Direction-aware system, motion tracking integration	PSNR, SSIM, temporal consistency
Qualitative	Visual motion and view effects	Dynamic scene realism
Future Work	Transition to Gaussian Splatting	Rendering quality, efficiency

D. Evaluation

My evaluation metrics will include both quantitative measures (PSNR, SSIM, temporal consistency scores) and qualitative assessments of motion handling and view-dependent effects. We anticipate that initial results from these experiments will provide valuable insights for further optimization and refinement of the integration approach.

IV. FUTURE DIRECTION

Looking ahead, I am particularly excited about transitioning this enhanced approach to Gaussian Splatting. Given Gaussian Splatting's superior performance in terms of rendering quality and training efficiency, we believe our direction and motion-aware enhancements could yield even more impressive results in this context. While this transition will present its own technical challenges - particularly in adapting our motion tracking and directional encoding to work with Gaussian primitives - the potential benefits in terms of rendering quality and computational efficiency make this a compelling next step. The lessons learned from our NeRF implementation will provide crucial insights for this future adaptation, potentially leading to even more efficient and effective dynamic scene representation methods.

REFERENCES

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf," *Communications of the ACM*, vol. 65, no. 1, p. 99–106, Dec 2021.

[2] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-nerf: Neural radiance fields for dynamic scenes," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 10313–10322, Jun 2021.

[3] S. Rojas, J. Zarzar, J. C. Pérez, A. Sanakoyeu, A. Thabet, A. Pumarola, and B. Ghanem, "Re-rend: Real-time rendering of nerfs across devices," *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, p. 3609–3618, Oct 2023.

[4] K. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla, "Nerfies: Deformable neural radiance fields," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, p. 5845–5854, Oct 2021.

[5] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, p. 1–14, Jul 2023.

[6] Z. Guo, W. Zhou, L. Li, M. Wang, and H. Li, "Motion-aware 3d gaussian splatting for efficient dynamic scene reconstruction," 2024. [Online]. Available: https://arxiv.org/abs/2403.11447

[7] A. Lou, B. Planche, Z. Gao, Y. Li, T. Luan, H. Ding, M. Zheng, T. Chen, Z. Wu, and J. Noble, "Dareplane: Direction-aware representations for dynamic scene reconstruction," 2024. [Online]. Available: https://arxiv.org/abs/2410.14169

[8] 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. [Online]. Available: https://arxiv.org/abs/2103.02597

[9] A. Lou, X. Yao, Z. Liu, J. Han, and J. H. Noble, "Self-supervised surgical instrument 3d reconstruction from a single camera image," *Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling*, p. 14, Apr 2023.