REAL-TIME MOTION CONSISTENCY RENDERING IN 4D GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review



Figure 1: We present Motion 4D Gaussian Splatting, a novel approach for real-time rendering of dynamic scenes at high resolutions. The proposed approach surpasses previous methods in rendering quality while reducing training time and improving inference speed on the Plenoptic Dataset Li et al. (2022b). In the right figure, the radius of the point represents the duration of the training.

ABSTRACT

Novel view synthesis for dynamic scenes is a key challenge in computer vision, due to the complexities of the time dimension and diverse motion patterns. The added dimension in 4D Gaussian distributions complicates their constraints, hindering the rendering of realistic and consistent appearances. To achieve realistic dynamic appearance reconstruction, we introduce Motion 4D Gaussian Splatting (Motion4DGS), a rendering approach that integrates motion consistency regularization into 4DGS. This regularization incorporates both 4D spatial and temporal consistency, leveraging depth and optical flow information. Furthermore, we extend optical flow constraints to continuous unit time intervals, enhancing the consistency of motion appearance without increasing the point cloud size. Motion4DGS excels in rendering dynamic scenes with multiple objects, including those with complex reflective areas. Comprehensive evaluations on both monocular and multi-view datasets show that Motion4DGS achieves real-time performance at high resolutions, surpassing existing methods in both quantitative and qualitative metrics. It provides a notable improvement in quality, all while maintaining comparable rendering speeds and training times to previous approaches.

INTRODUCTION

3D reconstruction is essential across various industries. While significant progress has been made in static scene synthesis Mildenhall et al. (2020); Knapitsch et al. (2017); Hedman et al. (2018); Barron et al. (2022), dynamic scenes Pumarola et al. (2021); Park et al. (2021b); Li et al. (2022b); Wu et al. (2020); Cheng et al. (2023) present a greater challenge due to the unpredictability of temporal dynamics and varied motion patterns. A key difficulty is ensuring smooth transitions between frames in dynamic sequences. In static scenes, the primary goal is accurate representation and rendering, but dynamic scenes also require maintaining the continuity and coherence of motion over time. Subtle

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

variations in texture, shading, and lighting need to be captured and rendered to preserve visual realism and physical plausibility. These changes, such as moving surfaces, need to be modeled to ensure they appear consistent and lifelike across frames, without abrupt transitions or unnatural artifacts.

To address the challenges of dynamic view synthesis, various methods have been proposed Du et al. (2021); Gao et al. (2021). NeRF Mildenhall et al. (2020) leverages volumetric rendering to generate 2D images from 3D data Park et al. (2021a;b); Tretschk et al. (2021); Pumarola et al. (2021); Fang et al. (2022); Li et al. (2022b); Song et al. (2023), but it is computationally demanding. While NeRF delivers impressive results, its dependence on numerous ray samples leads to slow rendering times, making it unsuitable for real-time applications, especially in dynamic environments. On the other hand, 3DGS Kerbl et al. (2023) enhances rendering speed by representing scenes with 3D Gaussian distributions, enabling real-time performance. Despite these advantages, 3DGS faces challenges in realistic rendering, and these issues are exacerbated in 4D space, where the introduction of the time dimension complicates the task of maintaining consistency across the 4D Gaussians. Temporal inconsistency disrupts the realism and visual coherence of dynamic scene rendering, leading to unrealistic and fragmented appearances.

Expanding 3D to 4D representations for dynamic scenes introduces the challenge of **ensuring Gaussian consistency** over time. As Gaussians evolve, they result in jittery or inconsistent motion. Achieving consistent rendering of dynamic objects' appearances remains **an under-constrained problem**. **The inherently smooth nature** of Gaussian distributions makes it difficult to capture fine details from sparse inputs in dynamic scenes, often leading to blurry surfaces. Moreover, **the absence of explicit constraints** linking Gaussian distributions to realistic visual properties leads to inconsistencies in both geometry and lighting. This issue is especially noticeable in moving object surfaces, where contours blur and appearance patterns misalign. While increasing the number of Gaussian primitives can improve accuracy, it also escalates computational and storage requirements, creating bottlenecks in real-time or resource-constrained settings, and complicating the trade-off between achieving realistic visuals and maintaining training costs.

To guide Gaussian distributions for consistent appearance in moving objects, we decouple the 4D spatiotemporal space into spatial and temporal components. We first apply a 4D spatial consistency regularization based on depth information, which aligns Gaussian distributions with real-world 3D structures, ensuring spatial coherence. This regularization maintains accurate motion scale and direction, especially in scenes with varying depth layers. In dynamic scenes with fast-moving objects, it enables Motion4DGS to capture reliable spatial information, rendering consistent 3D structures even under rapid motion, effectively handling challenging dynamic video sequences. Moreover, sparse input from real-world dynamic scenes can introduce uncertainty in 3D rendering. Motion4DGS tackles this by combining depth information with Gaussian distributions, resulting in robust renderings through smooth depth priors. The 4D spatial consistency regularization achieves high-fidelity rendering without increasing the oversize of the point cloud, avoiding the computational overhead of densification, even in noisy and sparsely observed scenes.

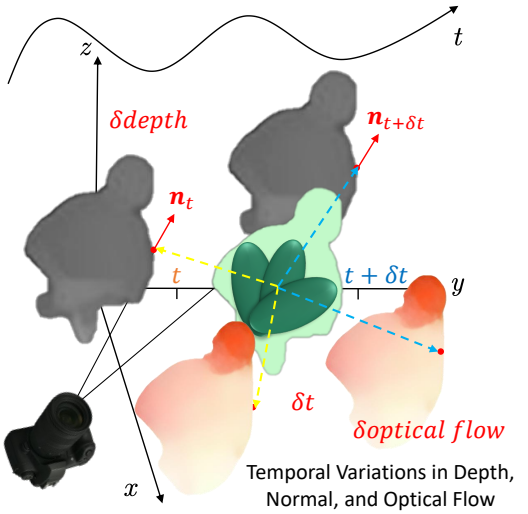

Figure 2: **Simplified Illustration.** We impose constraints on depth, surface normals, and optical flow over continuous unit time intervals rather than relying solely on discrete time steps, effectively addressing motion inconsistencies in dynamic scenes.

Furthermore, we integrate optical flow with spatial consistency to form 4D temporal consistency regularization. Depth alone cannot fully capture motion, but combining it with optical flow ensures appearance fidelity and temporal coherence. Optical flow refines Gaussian distribution depth and position, improving the handling of occlusion, motion blur, and rapid scene changes. The 4D tem-

poral consistency regularization uses optical flow to guide the evolution of Gaussians over time, ensuring their motion aligns with the scene's dynamics and adheres to physical laws. With accurate depth and optical flow, our 4D temporal consistency regularization enables Motion4DGS to predict object motion between frames, filling gaps in sparse or fast-moving scenes, which is essential for real-time and sparse video data. It efficiently constrains motion and deformation of Gaussian primitives across time, minimizing computational and memory overhead for complex dynamic scenes. Leveraging complementary temporal priors, our approach enhances appearance rendering by jointly regularizing space and time, allowing Motion4DGS to accurately predict motion and deliver high-quality, realistic dynamic rendering.

However, constraining depth, surface normals, and optical flow at a single time step is insufficient. To further address **the inconsistency of Gaussians representing the same feature** across different time steps, we extend motion consistency regularization to **the unit time interval** (Fig. 2). This is especially important for 4D Gaussians, which are harder to maintain consistently due to their higher dimensionality. By leveraging the unit time interval to constrain depth, surface normals, and optical flow, our approach resolves inconsistencies in Gaussian representations and learns transferable, invariant motion features, improving the subtlety and dynamic quality of the rendered object's appearance. Furthermore, dynamic features' change over the unit time interval simulates appearance variations from reflection and refraction on complex surfaces, enhancing realism. By enforcing motion features consistency across time, Motion4DGS learns consistent appearance representations, including geometry and lighting in dynamic scenes. This boosts rendering accuracy and lighting realism, allowing Motion4DGS to precisely render appearances of moving objects, even with sparse input data. Our motion consistency regularization ensures that Gaussian distributions follow smooth, physically plausible motion trajectories, maintaining both spatial and temporal coherence. This guarantees that the object's motion aligns with the scene's overall dynamics, resulting in realistic and consistent movement. Motion4DGS effectively manages non-rigid deformations, fluid-like dynamics, and interactions between complex objects, providing realistic motion rendering in challenging dynamic scenes (Fig. 1). In summary, our contributions are as follows:

- The introduction of **motion consistency regularization**, incorporating both 4D temporal and spatial consistency, effectively addresses the challenge of **rendering realistic dynamic appearances**.
- The first algorithm extends **the constraints of depth, surface normals, and optical flow** from separate time steps to **unit time intervals**, ensuring that 4D Gaussian motion remains consistent and smooth over time.
- An optimization method for 4D Gaussian properties, integrated **spatial and temporal information** to dynamically refine Gaussian distributions, effectively learning **consistent appearance representations**.
- A fast, differentiable **4D rendering approach** that achieves **real-time, realistic dynamic appearance rendering** while maintaining competitive FPS and training time.

Related Work

Gaussian-Driven NVS Novel view synthesis is a core challenge in 3D rendering Broxton et al. (2020); Collet et al. (2015); Li et al. (2018); Guo et al. (2015); Su et al. (2020); Guo et al. (2019); Hu et al. (2022); Li et al. (2017); Flynn et al. (2019). Conventional methods typically use interpolation or geometric models to represent scenes Levoy & Hanrahan (1996); Gortler et al. (1996); Debevec et al. (1996); Riegler & Koltun (2020); Thies et al. (2019); Waechter et al. (2014); Wood et al. (2023); Kutulakos & Seitz (2000); Penner & Zhang (2017); Seitz & Dyer (1999); Mildenhall et al. (2019); Srinivasan et al. (2019); Zhou et al. (2018). While NeRF-based methods rely on implicit radiance fields for scene representation Mildenhall et al. (2021); Barron et al. (2021); Zhang et al. (2020); Pumarola et al. (2021); Park et al. (2021a;b), subsequent approaches have explored more structured representations, including voxel grids Fang et al. (2022); Yi et al. (2023); Guo et al. (2023), factorized spatiotemporal feature grids Cao & Johnson (2023); Fridovich-Keil et al. (2023); Shao et al. (2023), masked space-time hash encodings Wang et al. (2023a), and other voxel-based formulations Gan et al. (2023), while incorporating motion cues such as optical flow to better capture scene dynamics Gao et al. (2021); Tian et al. (2023); Zhou et al. (2023). Despite significant progress Verbin et al. (2022); Kopanas et al. (2022); Bemana et al. (2022); Yan et al. (2023); Müller

et al. (2022); Fridovich-Keil et al. (2022), achieving real-time rendering for dynamic scenes, especially with monocular input, continues to be a challenging problem Wang et al. (2023c); Peng et al. (2023); Xu et al. (2022b); Lin et al. (2023); Wang et al. (2023b); Mildenhall et al. (2020); Chen et al. (2022). 3D Gaussian Splatting (3DGS) Kerbl et al. (2023) offers a solution by modeling complex geometries with Gaussian distributions, allowing for efficient real-time rendering with GPU acceleration Zhang et al. (2022); Xu et al. (2022a); Abou-Chakra et al. (2022), even when temporal dimensions are involved Yang et al. (2023b); Wu et al. (2023); Yang et al. (2023a); Duan et al. (2024). However, the inclusion of the time dimension complicates the constraints on Gaussians, leading to geometric degradation, noise, and artifacts. To address moving appearance inconsistencies, our motion consistency regularization includes 4D space-time regularization based on depth and optical flow. By extending temporal regularization across unit time intervals, Motion4DGS ensures precise, consistent appearance rendering in dynamic scenes.

NVS for Dynamic Scenes. Generating novel views of dynamic scenes from 2D images presents a significant challenge due to the complexities of temporal dependencies. Traditional methods often use geometric structures to capture dynamic scene elements Li et al. (2012); Collet et al. (2015); Kanade et al. (1997); Zitnick et al. (2004). Recent methods either condition neural scene representations on temporal inputs Du et al. (2021), or learn deformation fields that warp points from observed frames into a canonical radiance field Tretschk et al. (2021). Other approaches capture scene dynamics through complementary strategies, including scene decomposition Song et al. (2023), keyframe-initialized hierarchical training Li et al. (2022b), keyframe-based dynamic volumes Attal et al. (2023), flow-based motion modeling Li et al. (2021), explicit grid-based streaming radiance fields Li et al. (2022a) and time-aware neural voxel representations Fang et al. (2022). However, NeRF-based methods typically involve high computational and storage costs. In contrast, 3DGS-based methods take advantage of GPU acceleration, enabling real-time rendering of dynamic scenes, with some leveraging topological priors from deformation fields to enhance performance Yang et al. (2023b); Huang et al. (2024); Wu et al. (2023). However, geometric priors alone are inadequate for learning consistent representations of moving objects. To address the challenge of motion inconsistency, we integrate depth, surface normals, and optical flow as a temporal prior, effectively reducing motion blur and flickering artifacts. Furthermore, within the unit time interval, by jointly optimizing for spatial geometry and temporal dynamics, Motion4DGS achieves physically consistent and high-fidelity rendering of complex dynamic scenes, even in the presence of non-rigid deformations and sparse observations, maintaining real-time performance without increasing the magnitude of point counts.

METHOD

MOTION 4D GAUSSIAN SPLATTING FRAMEWORK

The overall framework of Motion4DGS is illustrated in Fig. 3. Initially, we use HexPlane Fridovich-Keil et al. (2023) to decouple spatial and temporal features by breaking the information into six 2D planes. These features are then combined into full 4D Gaussian spatiotemporal representations via a lightweight multi-layer perceptron (MLP). This combined feature is subsequently fed into a Gaussian deformation field network, which predicts the deformation of the 4D Gaussians for the next time step. Unlike previous methods, Motion4DGS incorporates the unit time interval δt in the input, enabling the deformation network to predict new Gaussian attributes, i.e., depth change $\delta depth$. Our motion consistency regularization, based on this interval, improves the accuracy and realism of dynamic scene rendering. By constraining depth and optical flow changes across objects within the interval, we effectively address the challenge of real-time rendering of realistic dynamic appearances.

In 4D spacetime, decoupling time and space with learnable time-conditioned parameters can cause the model to produce varying 3D Gaussian distributions at different time points, disrupting spatiotemporal consistency. Additionally, while 4D-Rotor GS separates time and space using quaternions, it struggles to integrate spatial priors from previous time steps, leading to poor generalization and performance in real-time dynamic rendering. To ensure spatiotemporal consistency in 4DGS, we learn the Gaussian distribution in standard space and use a deformation field to predict position and shape changes for the next time step. Motion4DGS allows Gaussian distributions to model dynamic scenes while preserving spatial priors, maintaining spatial consistency.

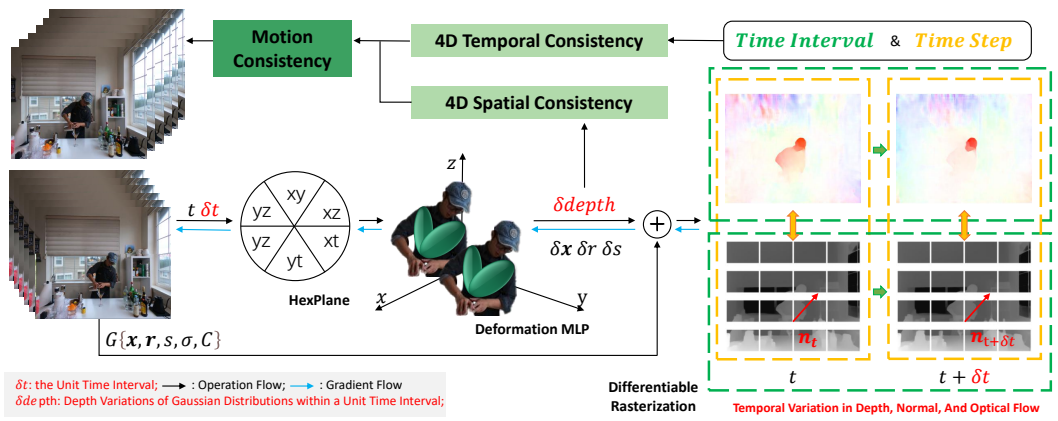

Figure 3: **Framework Overview.** We initialize 3D Gaussians from a sparse SfM point cloud and encode spatiotemporal features with depth variation $\delta depth$, which are rendered using differentiable rasterization. Motion consistency is enforced through complementary 4D spatial and 4D temporal regularizers on depth, surface normals, and optical flow across continuous unit time intervals.

However, relying only on 3D positions is inadequate for maintaining spatial consistency in dynamic scenes. To address this, we integrate depth information and extend spatial priors to create 4D spatial consistency regularization. Our approach ensures nearby Gaussian primitives with similar depths maintain consistent positions and motion, reducing unnatural spatial changes and producing smoother, coherent geometry. Motion4DGS accurately handles occlusion, preserving the hierarchical structure of scene elements. Moreover, temporal consistency is crucial for realistic rendering in dynamic scenes. To achieve this, we integrate optical flow into 4D temporal consistency regularization. This constrains the speed and direction of Gaussian distributions, ensuring smooth motion and improving the efficiency and realism of dynamic scene rendering. To further resolve inconsistencies in Gaussian representations across time steps, we introduce unit time intervals for spatiotemporal modeling. By constraining the depth, surface normals, and optical flow of the Gaussian distribution within these intervals, we effectively address the motion inconsistency in dynamic scenes. We quantify temporal changes by the unit time interval and use depth and optical flow to learn accurate motion representations for dynamic scenes. As the unit time interval approaches zero, the predicted motion closely matches the true motion.

Optimization Scheme

We now provide a detailed explanation of our motion consistency regularization scheme. Our motion consistency regularization combines 4D spatial and temporal consistency. Depth is crucial for smooth motion in dynamic scenes, providing spatial information that ensures consistent motion trajectories. In 4D spacetime, spatial consistency regularization uses depth information to accurately define object boundaries, preserving spatial distinctions across layers. Aligning Gaussian distributions with the correct depth creates realistic motion patterns, where objects closer to the camera move faster, and those farther away move slower. This prevents motion overlap between the foreground and background. Moreover, by incorporating spatial context, abrupt and jittery movements are minimized, ensuring smooth and continuous object trajectories over time. The 4D spatial consistency regularization aligns the Gaussian distribution with real-world 3D structures through geometric constraints, enhancing the stability, consistency, and physical accuracy of dynamic scene rendering. Given that $\boldsymbol{normal}$ represents the surface normal, and $\nabla \boldsymbol{D}_i(t)$ denotes the gradient of the depth map at the i-th pixel in the ground truth, the 4D spatial consistency regularization is defined as:

$$\mathcal{L}_{4DSpace} = \sum_{i=1}^{N} \left\| \boldsymbol{normal}_i(t) - \nabla \boldsymbol{D}_i(t) \right\|^2. \tag{1}$$

where the surface normal is computed from the depth map.

Object motion requires both depth-based spatial context and precise direction and magnitude. To refine motion representation, we incorporate optical flow. In dynamic scenes, motion inconsistencies

reduce realism across time steps. Optical flow ensures that Gaussian distributions follow physical laws between time steps. Our 4D temporal consistency regularization, built on optical flow, maintains consistent motion and prevents misalignment. Integrating optical flow into Motion4DGS addresses motion blur and flickering artifacts while guiding Gaussian distributions to evolve in sync with the scene's movement, ensuring physical accuracy. Define $\boldsymbol{flow}_i(t)$ as the optical flow at the i-th pixel and θ as the threshold for motion velocity. The 4D temporal consistency regularization for a single time step is given by the following expression:

$$\mathcal{L}_{TimeStep} = \sum_i |\boldsymbol{flow}_i(t) \cdot \boldsymbol{D}_i(t)|^2 \cdot \mathbb{I}(|\boldsymbol{flow}_i(t)| > \theta). \tag{2}$$

When optical flow exceeds a certain threshold, we use the velocity and directional temporal priors from the flow to guide smooth variations in the motion features of the Gaussian distribution, including depth. By incorporating accurate temporal priors, our 4D temporal consistency regularization helps guide the evolution of dynamic scenes and enables future motion prediction for smooth time transitions. We ensure that the motion of Gaussian distributions adheres to physical laws across time steps, maintaining consistent motion and preventing misalignment. Ultimately, Motion4DGS ensures both the smoothness and physical accuracy of dynamic scene rendering, especially for complex motions and deformations.

However, constraining the depth and motion flow of Gaussian distributions at a single time step is insufficient. Our motion consistency regularization further aligns depth, surface normals, and optical flow from separate time steps to continuous unit time intervals. Depth variations over time are more crucial than at a single step, as 4D Gaussians are harder to maintain consistently due to the temporal dimension. A key challenge in dynamic scenes is ensuring Gaussians representing the same object remain consistent across time, especially during frame transitions. Our regularization imposes constraints on depth, surface normals, and optical flow across time and space, preventing unrealistic jumps in depth maps and optical flow, ensuring smooth transitions and greater consistency in dynamic scenes. Given the unit time interval δt, we constrain depth and surface normals as:

$$\mathcal{L}_{TimeIntervalD} = \textstyle\sum_{\delta t} \sum_{i=1}^{N} \|\boldsymbol{normal}_i(t+\delta t) - \nabla \boldsymbol{D}_i(t+\delta t) - (\boldsymbol{normal}_i(t) - \nabla \boldsymbol{D}_i(t))\|^2. \tag{3}$$

Furthermore, motion blur in undersampled dynamic regions can be improved with smooth, consistent changes in optical flow. Our motion consistency regularization ensures smooth transitions of Gaussian distributions over time by applying depth-based spatial continuity and optical flow-based temporal continuity. This prevents underfitting in sparse and noisy regions, maintains consistent geometry without unrealistic depth and motion changes, avoiding the high cost of point cloud densification. By constraining Gaussian variations over the unit time interval, our approach enhances the detailed surface geometry of moving objects without increasing the overall point cloud size, improving both appearance and realism.

$$\mathcal{L}_{TimeIntervalF} = \sum_{\delta t} \sum_{i=1}^{N} |\boldsymbol{flow}_i(t+\delta t) - \boldsymbol{flow}_i(t)|^2 \cdot \mathbb{I}(\boldsymbol{D}_i(t) > 0). \tag{4}$$

In regions with consistent depth values, we leverage optical flow variations within a unit time interval to constrain the motion anomaly of the Gaussian distribution. This approach guides Motion4DGS in accurately predicting the motion trajectory for the next time step in dynamic scenes. Our regularization across continuous time intervals is represented as:

$$\mathcal{L}_{TimeInterval} = \mathcal{L}_{TimeIntervalD} + \lambda_1 \mathcal{L}_{TimeIntervalF}. \tag{5}$$

The final 4D temporal consistency regularization is defined as:

$$\mathcal{L}_{4DTime} = \mathcal{L}_{TimeInterval} + \lambda_2 \mathcal{L}_{TimeStep}. \tag{6}$$

Finally, we can obtain the total motion consistency loss:

$$\mathcal{L}_{Motion} = \mathcal{L}_{4DSpace} + \lambda_3 \mathcal{L}_{4DTime}. \tag{7}$$

where λ_1, λ_2, and λ_3 are learned hyperparameters.

EXPERIMENTS

DATASETS AND IMPLEMENTATION DETAILS

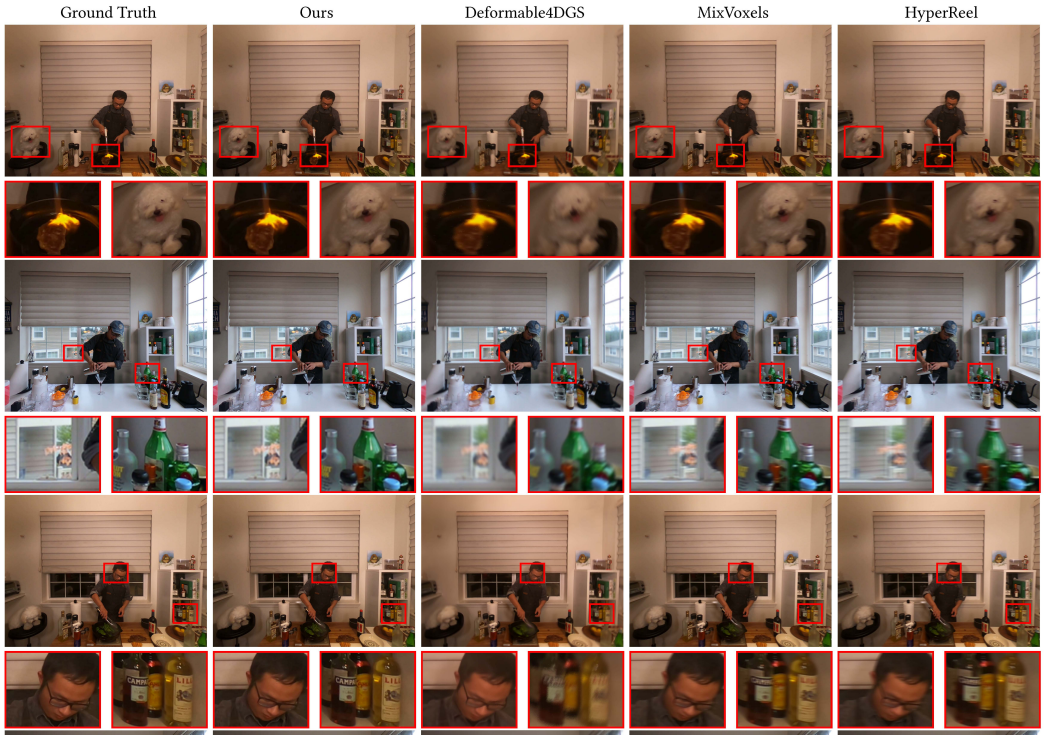

Figure 4: **Qualitative Comparison on Plenoptic Video Dataset**. Our approach surpasses strong baselines by capturing realistic details in dynamic regions, *e.g.*, the fast-moving kitchenware. Furthermore, it produces the sharpest reflection details, *e.g.*, the light reflections on glass bottles.

We evaluate our approach using two well-established datasets that encompass a broad range of challenges inherent to dynamic scene modeling. **Plenoptic Video Dataset Li et al. (2022b)** contains six different real-world scenes, each offering 17 to 20 views for training, while one central view is set aside for evaluation. All images have a resolution of 1352×1014 Li et al. (2022b). **D-NeRF Dataset Pumarola et al. (2021)** comprises monocular videos from 8 different scenes. Each scene contains 50 to 200 images for training, 10 to 20 images for validation, and 20 images for testing. All images have been resized to a standard resolution of 800×800 Pumarola et al. (2021).

Table 1: **Quantitative Comparison on Plenoptic Video Dataset**. We compare Motion4DGS with SOTA methods. Motion4DGS substantially outperforms the baselines in PSNR, achieving the shortest training times and comparative rendering speeds. *: trained on 8 GPUs and tested only on the Flame Salmon scene.

ID	Method	PSNR↑	SSIM↑	LPIPS↓	Train↓	FPS↑
a	DyNeRF Li et al. (2022b)*	29.58	-	0.08	1344 h	0.015
b	StreamRF Li et al. (2022a)	28.16	0.85	0.31	79 min	8.50
c	HyperReel Attal et al. (2023)	30.36	0.92	0.17	9 h	2.00
d	NeRFPlayer Song et al. (2023)	30.69	-	0.11	6 h	0.05
e	K-Planes Fridovich-Keil et al. (2023)	30.73	0.93	0.07	190 min	0.10
f	MixVoxels Wang et al. (2023b)	30.85	0.96	0.21	91 min	16.70
g	RealTime4DGS Yang et al. (2023a)	29.95	0.92	0.16	8 h	72.80
h	Deformable4DGS Wu et al. (2023)	28.42	0.92	0.17	72 min	39.93
i	Ours	33.20	0.96	0.10	20 min	110.00

Optical flow characterizes pixel motion over the unit time interval. We adopt standard optical-flow benchmarks for evaluation. In scenarios with camera rotation, accurate rendering of realistic dynamic scenes merely requires augmenting the model with camera motion. Our implementation is built primarily on the PyTorch Paszke (2019) framework, with all tests conducted on a single RTX 3090 GPU. We fine-tune the optimization parameters following the configuration settings used in 3DGS Kerbl et al. (2023).

RESULTS

Evaluation on Plenoptic Video Dataset. As shown in Tab. 1, earlier methods struggle to balance rendering speed and quality due to slow volumetric rendering (*a-f*), high computational costs (*c*, *d*, *g*), and inefficient modeling of dynamic motion (*a*, *c*, *g*). In contrast, Motion4DGS, achieves



Figure 5: **Qualitative Comparison on D-NeRF Dataset.** Our approach excels in rendering realistic details in various regions, *e.g.*, the complex geometries inside the excavator bucket with high precision in *Lego*.

substantial improvements in both speed and quality, outperforming existing techniques in rendering high-resolution videos (1352×1014) with superior fidelity. Motion4DGS also outperforms Deformable4DGS, improving PSNR by 3.36. As shown in Fig. 4, our approach produces the sharpest rendering of dynamic regions when compared to baseline methods. Motion4DGS excels at capturing high-frequency details in fast-moving areas, such as human fingers and kitchen utensils. Notably, Motion4DGS excels in rendering specular regions, with the brightness distribution of glass bottles and flames closely matching the ground truth. By incorporating motion consistency regularization, Motion4DGS effectively resolves temporal inconsistencies, resulting in highly realistic dynamic appearance rendering.

Evaluation on D-NeRF Dataset As shown in Tab. 2, Motion4DGS outperforms previous methods, increasing the PSNR of Deformable4DGS from 32.99 to 34.37. Fig. 5 highlights our approach effectively resolves appearance inconsistencies: *e.g.*, the *T-Rex* scene presents the shape edges of teeth, the *Hook* scene features clear finger contours and realistic glove textures, and the *Stand Up* scene captures real-world reflections with faithful brightness distribution. Notably, the new perspective in the *Lego* scene details texture patterns. Motion4DGS ensures that the motion of the objects in the scene is smooth and realistic, addressing dynamic appearance inconsistencies.

Table 2: **Quantitative Comparison on D-NeRF Dataset**. Motion4DGS outperforms both NeRF-based and Gaussian-based baselines, achieving higher PSNR and faster training times while producing high-quality renders.

Method	PSNR↑	SSIM↑	LPIPS↓	Train↓	FPS↑
D-NeRF Pumarola et al. (2021)	29.17	0.95	0.07	24 h	0.13
TiNeuVox Fang et al. (2022)	32.87	0.97	0.04	28 min	1.60
K-Planes Fridovich-Keil et al. (2023)	31.07	0.97	0.02	54 min	1.20
Deformable3DGS Yang et al. (2023b)	39.31	0.99	0.01	26 min	85.45
RealTime4DGS Yang et al. (2023a)	29.95	0.92	0.16	8 h	72.80
Deformable4DGS Wu et al. (2023)	32.99	0.97	0.05	13 min	104.00
Ours	35.30	0.99	0.02	5 min	220.00

Ablation Studies

4D Spatial Consistency. ' To assess the effectiveness of our spatial consistency regularization, we activate only this term in the ablation. As shown in Tab. 3 (*b*), this yields substantial gains in rendering accuracy and stability: PSNR increases by 2.32 dB, SSIM rises from 0.954 to 0.986, and LPIPS exhibits a relative decrease of 59.5%. Furthermore, as shown in Fig. 6, removing the space term weakens spa-

Table 3: **Ablation Study with Quantitative Comparison on D-NeRF Dataset**.

ID	Space	Motion	Time	PSNR↑	SSIM↑	LPIPS↓
a				32.73	0.954	0.084
b	✓			35.05	0.986	0.034
c		✓		33.60	0.970	0.060
d			✓	34.86	0.975	0.048
Full	✓	✓	✓	35.30	0.994	0.018

tial anchoring and causes local geometric and textural drift: edges on the blinds and frame soften with more fine-scale jaggies and misalignment; texture contrast on the countertop and bottle diminishes; and object boundaries loosen. Experimental results indicate that enforcing consistency over local spatial neighborhoods in the 4D setting stabilizes geometry and appearance. By refining surface normals and depth estimates, our approach effectively resolves motion-induced inconsistencies, suppresses high-frequency artifacts, and jointly improves perceptual quality.

Motion Consistency at Separate Time Steps. As shown in the Tab. 3 (*c*), enabling only the motion term separately yields consistent gains: PSNR increases from 32.73 to 33.60 dB, SSIM from 0.954 to 0.970, and LPIPS decreases by approximately 28.6%. Fig. 6 further shows that removing the motion term affects non-rigid regions. Moving objects show slight trailing and double edges, and occlusion boundaries are less clean, indicating that temporal correspondence and deformation estimates jitter without motion-consistency constraints. Experimental results indicate that our motion consistency constraint effectively suppresses inter-frame jitter and ghosting, with pronounced improvements in perceptual quality and structural consistency, especially in regions undergoing large non-rigid displacements.



Figure 6: **Qualitative Comparison of the Ablation Study.**

4D Temporal Consistency. We enhance temporal modeling by unit time interval consistency to optimize the rendering quality of dynamic scenes. As shown in Tab. 3 (*d*), temporal regularization enables substantial gains: PSNR increases by 2.13 dB, SSIM rises from 0.954 to 0.975, and LPIPS decreases by approximately 42.9%. Fig. 7 further illustrates how this temporal consistency constraint refines the rendering of fine dynamic details, *e.g.*, the clear outline of kitchen tongs. In Fig. 6, removing the time term degrades temporal stability: brightness and highlight intensity vary more, with mild exposure and hue shifts visible in bottle reflections and tabletop highlights. Experimental results indicate that temporal consistency constraints are pivotal in achieving high fidelity and low perceptual error. Combined with the space and motion terms, Motion4DGS further stabilizes fine details and suppresses residual flicker, enabling physical realism rendering.

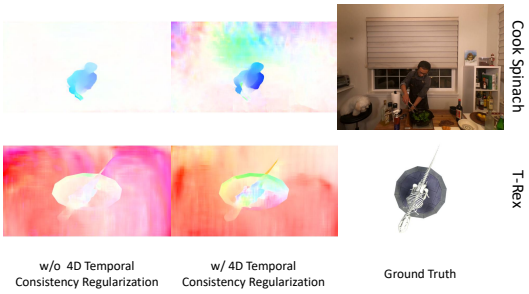

Figure 7: **Optical Flow Visualization.** We visualize the dynamics generated by Motion4DGS. By combining temporal regularization, Motion4DGS effectively renders realistic dynamic appearances, *e.g.*, the detailed edges of the dinosaur.

Conclusion and Future Works

We introduce Motion 4D Gaussian Splatting (Motion4DGS) for real-time, high-quality modeling of 4D dynamic scenes. To achieve realistic dynamic appearances, we implement motion consistency regularization, combining 4D spatial and temporal consistency using depth and optical flow. By extending motion consistency from individual time steps to continuous unit time intervals, we address Gaussian inconsistencies across time, improving dynamic scene appearance without expanding the point cloud size. Motion4DGS outperforms existing methods in rendering quality while maintaining competitive training times. In the future, we will integrate Motion4DGS into real-time dynamic scene generation systems, optimizing for VR/AR applications and exploring scalable solutions for large environments.

REFERENCES

Jad Abou-Chakra, Feras Dayoub, and Niko Sünderhauf. Particlenerf: Particle based encoding for online neural radiance fields in dynamic scenes. *arXiv preprint arXiv:2211.04041*, 2022.

Benjamin Attal, Jia-Bin Huang, Christian Richardt, Michael Zollhoefer, Johannes Kopf, Matthew O'Toole, and Changil Kim. Hyperreel: High-fidelity 6-dof video with ray-conditioned sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16610–16620, 2023.

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5855–5864, 2021.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5470–5479, 2022.

Mojtaba Bemana, Karol Myszkowski, Jeppe Revall Frisvad, Hans-Peter Seidel, and Tobias Ritschel. Eikonal fields for refractive novel-view synthesis. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–9, 2022.

Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, Peter Hedman, Matthew Duvall, Jason Dourgarian, Jay Busch, Matt Whalen, and Paul Debevec. Immersive light field video with a layered mesh representation. *ACM Transactions on Graphics (TOG)*, 39(4):86–1, 2020.

Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 130–141, 2023.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*, pp. 333–350. Springer, 2022.

Wei Cheng, Ruixiang Chen, Siming Fan, Wanqi Yin, Keyu Chen, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, et al. Dna-rendering: A diverse neural actor repository for high-fidelity human-centric rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19982–19993, 2023.

Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics (ToG)*, 34(4):1–13, 2015.

Paul E Debevec, Camillo J Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996.

Yilun Du, Yinan Zhang, Hong-Xing Yu, Joshua B Tenenbaum, and Jiajun Wu. Neural radiance flow for 4d view synthesis and video processing. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14304–14314. IEEE Computer Society, 2021.

Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: Towards efficient novel view synthesis for dynamic scenes. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2024.

Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–9, 2022.

John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. Deepview: View synthesis with learned gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2367–2376, 2019.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5501–5510, 2022.

Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12479–12488, 2023.

Wanshui Gan, Hongbin Xu, Yi Huang, Shifeng Chen, and Naoto Yokoya. V4d: Voxel for 4d novel view synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5712–5721, 2021.

Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996.

Kaiwen Guo, Feng Xu, Yangang Wang, Yebin Liu, and Qionghai Dai. Robust non-rigid motion tracking and surface reconstruction using l0 regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3083–3091, 2015.

Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)*, 38 (6):1–19, 2019.

Xiang Guo, Jiadai Sun, Yuchao Dai, Guanying Chen, Xiaoqing Ye, Xiao Tan, Errui Ding, Yumeng Zhang, and Jingdong Wang. Forward flow for novel view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16022–16033, 2023.

Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (ToG)*, 37(6):1–15, 2018.

Tao Hu, Tao Yu, Zerong Zheng, He Zhang, Yebin Liu, and Matthias Zwicker. Hvtr: Hybrid volumetric-textural rendering for human avatars. In *2022 International Conference on 3D Vision (3DV)*, pp. 197–208. IEEE, 2022.

Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4220–4230, 2024.

Takeo Kanade, Peter Rander, and PJ Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, 4(1):34–47, 1997.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4):1–14, 2023.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017.

Georgios Kopanas, Thomas Leimkühler, Gilles Rainer, Clément Jambon, and George Drettakis. Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (TOG)*, 41(6):1–15, 2022.

Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. *International journal of computer vision*, 38:199–218, 2000.

Marc Levoy and Pat Hanrahan. Light field rendering. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996.

Hao Li, Linjie Luo, Daniel Vlasic, Pieter Peers, Jovan Popović, Mark Pauly, and Szymon Rusinkiewicz. Temporally coherent completion of dynamic shapes. *ACM Transactions on Graphics (TOG)*, 31(1):1–11, 2012.

Lingzhi Li, Zhen Shen, Zhongshu Wang, Li Shen, and Ping Tan. Streaming radiance fields for 3d video synthesis. In *Advances in Neural Information Processing Systems*, pp. 13485–13498, 2022a.

Tianye Li, Mira Slavcheva, Michael Zollhoefer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, Richard Newcombe, et al. Neural 3d video synthesis from multi-view video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5521–5531, 2022b.

Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6498–6508, 2021.

Zhong Li, Yu Ji, Wei Yang, Jinwei Ye, and Jingyi Yu. Robust 3d human motion reconstruction via dynamic template construction. In *2017 International Conference on 3D Vision (3DV)*, pp. 496–505. IEEE, 2017.

Zhong Li, Minye Wu, Wangyiteng Zhou, and Jingyi Yu. 4d human body correspondences from panoramic depth maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2877–2886, 2018.

Haotong Lin, Sida Peng, Zhen Xu, Tao Xie, Xingyi He, Hujun Bao, and Xiaowei Zhou. High-fidelity and real-time novel view synthesis for dynamic scenes. In *SIGGRAPH Asia Conference Proceedings*, 2023.

B Mildenhall, PP Srinivasan, M Tancik, JT Barron, R Ramamoorthi, and R Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, 2020.

Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5865–5874, 2021a.

Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *ACM Transactions on Graphics (TOG)*, 40(6):1–12, 2021b.

A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

Sida Peng, Yunzhi Yan, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Representing volumetric videos as dynamic mlp maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4252–4262, 2023.

Eric Penner and Li Zhang. Soft 3d reconstruction for view synthesis. *ACM Transactions on Graphics (TOG)*, 36(6):1–11, 2017.

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10318–10327, 2021.

Gernot Riegler and Vladlen Koltun. Free view synthesis. In *European Conference on Computer Vision*, pp. 623–640, 2020.

Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35:151–173, 1999.

Ruizhi Shao, Zerong Zheng, Hanzhang Tu, Boning Liu, Hongwen Zhang, and Yebin Liu. Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16632–16642, 2023.

Liangchen Song, Anpei Chen, Zhong Li, Zhang Chen, Lele Chen, Junsong Yuan, Yi Xu, and Andreas Geiger. Nerfplayer: A streamable dynamic scene representation with decomposed neural radiance fields. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2732–2742, 2023.

Pratul P Srinivasan, Richard Tucker, Jonathan T Barron, Ravi Ramamoorthi, Ren Ng, and Noah Snavely. Pushing the boundaries of view extrapolation with multiplane images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 175–184, 2019.

Zhuo Su, Lan Xu, Zerong Zheng, Tao Yu, Yebin Liu, and Lu Fang. Robustfusion: Human volumetric capture with data-driven visual cues using a rgbd camera. In *The European Conference on Computer Vision*, pp. 246–264, 2020.

Justus Thies, Michael Zollhöfer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. *Acm Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

Fengrui Tian, Shaoyi Du, and Yueqi Duan. Mononerf: Learning a generalizable dynamic radiance field from monocular videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17903–17913, 2023.

Edith Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12959–12970, 2021.

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5481–5490, 2022.

Michael Waechter, Nils Moehrle, and Michael Goesele. Let there be color! large-scale texturing of 3d reconstructions. In *European Conference on Computer Vision*, pp. 836–850. Springer, 2014.

Feng Wang, Zilong Chen, Guokang Wang, Yafei Song, and Huaping Liu. Masked space-time hash encoding for efficient dynamic scene reconstruction. In *Advances in neural information processing systems*, pp. 70497–70510, 2023a.

Feng Wang, Sinan Tan, Xinghang Li, Zeyue Tian, Yafei Song, and Huaping Liu. Mixed neural voxels for fast multi-view video synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19706–19716, 2023b.

Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3295–3306, 2023c.

Daniel N Wood, Daniel I Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H Salesin, and Werner Stuetzle. Surface light fields for 3d photography. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 487–496. 2023.

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. *arXiv preprint arXiv:2310.08528*, 2023.

Minye Wu, Yuehao Wang, Qiang Hu, and Jingyi Yu. Multi-view neural human rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1682–1691, 2020.

Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5438–5448, 2022a.

Qingshan Xu, Weihang Kong, Wenbing Tao, and Marc Pollefeys. Multi-scale geometric consistency guided and planar prior assisted multi-view stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4945–4963, 2022b.

Zhiwen Yan, Chen Li, and Gim Hee Lee. Nerf-ds: Neural radiance fields for dynamic specular objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8285–8295, 2023.

Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *arXiv preprint arXiv:2310.10642*, 2023a.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. *arXiv preprint arXiv:2309.13101*, 2023b.

Taoran Yi, Jiemin Fang, Xinggang Wang, and Wenyu Liu. Generalizable neural voxels for fast human radiance fields. *arXiv preprint arXiv:2303.15387*, 2023.

Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.

Qiang Zhang, Seung-Hwan Baek, Szymon Rusinkiewicz, and Felix Heide. Differentiable point-based radiance fields for efficient view synthesis. In *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–12, 2022.

Kaichen Zhou, Jia-Xing Zhong, Sangyun Shin, Kai Lu, Yiyuan Yang, Andrew Markham, and Niki Trigoni. Dynpoint: Dynamic neural point for view synthesis. In *Advances in Neural Information Processing Systems*, pp. 69532–69545, 2023.

Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification. *ACM Transactions on Graphics*, 37(4):1–12, 2018.

C Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. *ACM transactions on graphics (TOG)*, 23(3):600–608, 2004.