ADVANCEMENTS IN RADIANCE FIELD TECHNIQUES FOR VOLUMETRIC VIDEO GENERATION: A TECHNICAL OVERVIEW

Joshua Maraval[1,2], Nicolas Ramin[1] and Lu Zhang[1,2]

[1] IRT b<>com, France and [2] CNRS, IETR-UMR 6164, France

ABSTRACT

Over the past decades, video consumption and video devices have become widespread globally. In 2014, mainstream virtual reality headsets marked a pivotal moment for 360° video accessibility. Advanced immersive devices, like the Apple Vision Pro as well as smartphones and tablets with advanced spatial capabilities can now provide users with real-time 6 Degrees of Freedom (6DoF) navigation experiences.

However, the lack of engaging content is hindering potential applications in areas such as training and entertainment. Volumetric video is a promising solution. However, its production poses challenges, such as the need for natural 3D+t reconstruction, coding, and rendering, which still require intensive computational resources.

In 2020, the ground-breaking Neural Radiance Field (NeRF) paper introduced a new way to generate natural free-viewpoint renderings of real scenes from sparsely captured views. Follow-up research has led to faster and more flexible methods, such as the widely used 3D Gaussian Splatting. However, these approaches require independent models for each frame, posing a challenge for volumetric video representation. To address temporal limitations, extensions of radiance field techniques use temporal redundancy to create a compact, temporally consistent, and editable volumetric video representation.

This paper offers a comprehensive overview of state-of-the-art volumetric video methods based on neural radiance fields, including their respective advantages and drawbacks. Using a diverse multi-view video dataset of diverse real-world scenarios, we present an objective evaluation of these methods for video volumetric content generation in entertainment and training.

INTRODUCTION

Novel view synthesis (NVS) is a long-standing challenge of 3D computer vision: the rendering of unseen views of a scene from a set of captured views. NVS has a growing impact on a wide array of video applications including media consumption [1], sports retransmission [2], immersive training [3] and telepresence [4]. The applications fall into one of two categories: visual effects or immersive experiences. One common visual effect with NVS is the virtual rendering of non-captured camera movement. An illustrative is the Intel True View technology [5], which proposes frozen time 360 degree replays of sports

stadiums. In contrast, immersive experiences rely on real-time NVS to display position-dependant views to a user, allowing them to navigate freely within a virtual scene as if they were in the real location.

Early NVS methods interpolated viewpoints from depth information [6]. These methods were capable of rendering realistic novel views in ideal conditions, but they had limited light effect rendering capacity and were restricted to rendering views that were close to the reference views. Concurrently, novel devices, including the Apple Vision Pro, were making real-time 6 Degrees of Freedom (6DoF) tracking increasingly accessible. Free navigation of real scenes requires reconstructing a complete representation of the scene from recorded videos. In recent years, significant progress has been made towards volumetric-based approaches. One such approach is Neural Radiance Fields (NeRF) [7] which was first published in 2020. This method enables the generation of high-quality views by modelling the scene's geometry and radiance. NeRF demonstrated the capacity of radiance field methods to represent complex real scenes with accurate light effects. Following the publication of NeRF, radiance fields have rapidly become the most regarded approach for NVS of natural content.

Early radiance field methods, including NeRF necessitated slow reconstructions for every scene to be reconstructed, and could not render novel views in real time. New approaches have been implemented from NeRF for faster processing [8], [9], higher quality rendering [10], [11] and more stable reconstructions [12]. The recent radiance field method 3D Gaussian Splatting techniques [13] has gained considerable popularity due to its demonstration of state-of-the-art rendering quality with in-real-time rendering capacity.

Building up upon the latest advances in radiance field methods, the volumetric video field has undergone a rapid evolution in recent years. New techniques have extended the applicability of radiance field to a range of classical 2D video tasks, including semantic segmentation [14], streaming [15] and edition [16]. The development of real-time dynamic representations of radiance fields has opened the door to 6DoF+t navigable content in real-time.

However, each approach has its own advantages and drawbacks. We propose an overview of NVS methods, focusing on radiance field approaches. We review the latest advances towards to real-time 6DoF+t navigation and evaluate the state-of-the-art methods on a dataset of scenes showcasing complex human interactions in diverse environments.

NOVEL VIEW SYNTHESIS METHODS

From multiple images capturing a single scene, NVS algorithms render novel viewpoints that have not been previously observed. In order to interpret the image's information, it is necessary to understand the position and orientation of the camera relative to the other images. There exist methods that generate novel views only from input images without requiring prior knowledge of camera parameters, simultaneously based on SLAM [17], [18] or state-of-the-art radiance fields [19], [20], [21]. These camera-parameters free methods are considered out of the scope of this review. This review focuses solely on methods that use known camera parameters of input images. In order to obtain camera parameters from multi-view images of a scene, a pre-processing step of structure-from-motion is typically used. In this paper, this calibration step is achieved using the structure-from-motion software COLMAP [18], as illustrated in Figure 1.

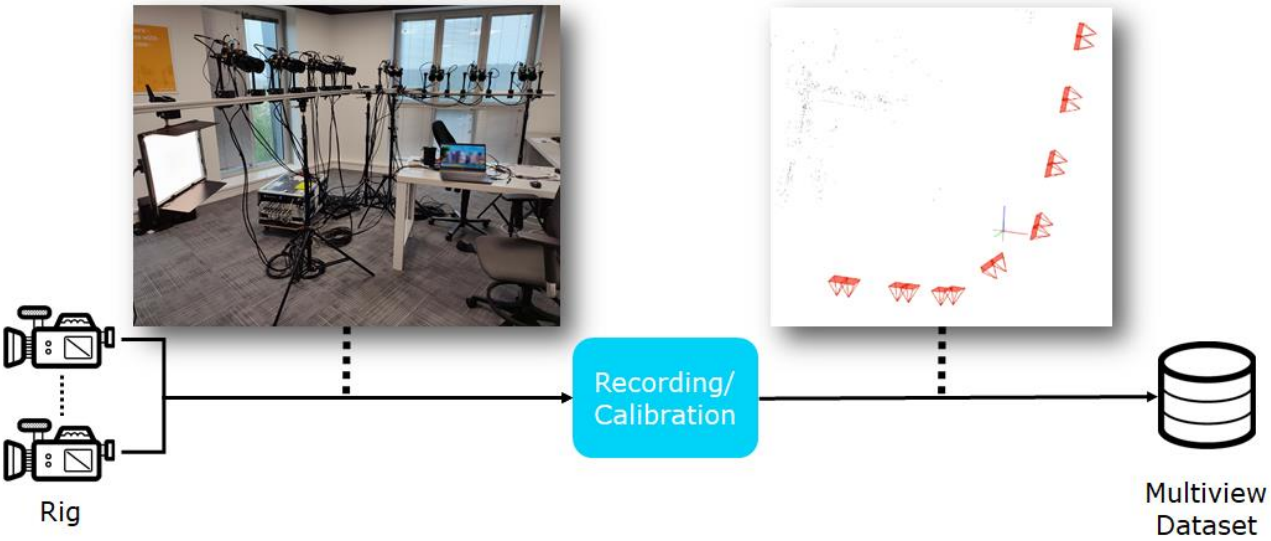

Figure 1 - The input images for NVS undergo a pre-processing calibration step. The camera parameters are retrieved from the input images with the Structure-from-motion software COLMAP. A Multiview Dataset is constituted of the input images and meta-parameters of the scene, including camera parameters.

Interpolation-based View Synthesis

Interpolation-based NVS methods generate novel views by interpolating pixel information between input views (11). Some approaches, called Depth Image Based Rendering (DIBR) leverage information from the associated depth maps of the input images for more accurate translation of the input pixels to the novel view [6], [22]. DIBR can achieve high-quality rendering of intermediate views, but often results in errors at object edges and occluded areas and lacks light effect rendering. Light field approaches [23], [24] interpolate all pixels in an intermediate 3D space, which is then inferred to render novel views. Light fields can render views with complex occlusions and light effects but this necessitates a dense array of input views.

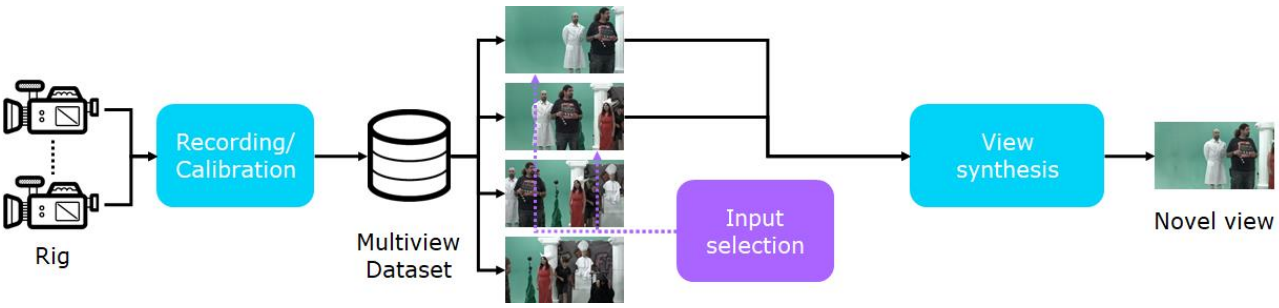

Figure 2 - Interpolation-based view synthesis. The pixels from the input views are interpolated to render the novel view.

Following the democratization of Convolutional Neural Networks (CNNs) based deep learning by Krizhevsky et al. in 2012 [25], CNNs gained popularity for view synthesis methods. In recent works, CNNs have been trained as a post-filter to improve the rendered images of DIBR-based methods, effectively removing some artefacts [26]. Other methods train a CNN-based architecture in place of the interpolation function for DIBR [27]. Generative Adversarial Networks (GANs) have demonstrated that a few images of a scene can be used to predict other views [28], [29] or enhance NVS renders [11]. While methods

based on deep CNNs for NVS may generate high-quality renders, they are inherently slow to infer which results in low rendering frame rates.

Learning-based volumetric representations

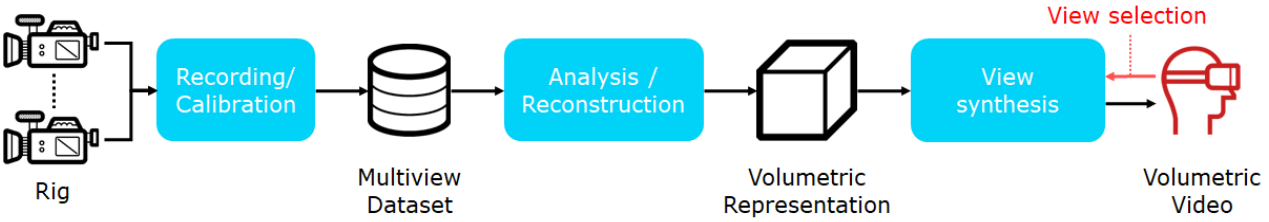

Figure 3 - 3D Model learning-based View synthesis from volumetric representation. A volumetric representation of the scene is reconstructed from the input images. After complete training, novel views are rendered by inference of the volumetric representation.

The rendering of views from a volumetric representation is a well-studied subject of computer vision. For instance, photorealistic models can be rendered from synthetic scenes modelled as meshes using the latest rendering technologies. However, models used to design synthetic data are limited representations of the real world and differ significantly from the light physics behind the human vision. Some works have proposed the use of CNN architectures for higher quality mesh rendering [30], [31], [32]. However, these representations are partially differentiable, which makes them difficult to optimize without a high density of input images. Seminal works proposed the use of a plenoptic function as a volumetric representation of scene, capable of rendering light coherent novel views [23], [24]. The plenoptic function is a 5D function describing the light flow at any 3D position of any 2D orientation. With recent advances in machine learning, a learning-based approach has become a viable option.

Most modern approaches to learning-based volumetric representations feature a complete or partial representation of the plenoptic function. A generic workflow is presented in Figure 3. Prior to rendering novel views, the volumetric representation must be reconstructed from the input images. Input views are rendered from the representation, and then an error loss is computed based on the difference between the rendered image and the reference images. The loss is then propagated backwards to adjust the volumetric model parameters into a new model that more closely renders the reference images, until complete convergence is achieved.

At the condition of having a fully differentiable rendering pipeline, a variety of volumetric representation can be trained to render novel views of a scene. Multi-plane images (MPI) approaches divide the plenoptic function into successive planes that store colour and transparency information [33], [34], [35]. While these approaches are fast and capable of photorealistic renderings, MPI are limited to rendering views facing a single direction, as a perpendicular view to the planes cannot be rendered. Broxton et al. [36] demonstrated that a structure of spherical planes can be used for efficient 360° scene rendering, with the limitation that rendered views must be close to the circle's centre.

NEURAL RADIANCE FIELDS

NeRF

The Neural Radiance Fields (NeRF) model, published in 2020 by Mildenhall et al. [7], marked a significant shift in the field of volumetric rendering. The paper attracted

considerable and growing attention, as evidenced by the citations graph in Figure 4. NeRF introduces a fully connected deep network that outputs volume density and view-dependent radiance at any point in space. A ray-casting strategy is proposed to retrieve the colours of any view. The pixel ray is projected into the three-dimensional space, sampled into three-dimensional points, and the density and radiance of each point are inferred by a multilayer perceptron (MLP). The pixel colour is obtained with a classical ray-casting rendering. The MLP is trained from scratch for any new scene reconstruction. NeRF is a powerful representation capable of generating photorealistic renders of complex scenes and a flexible and simple volumetric representation.

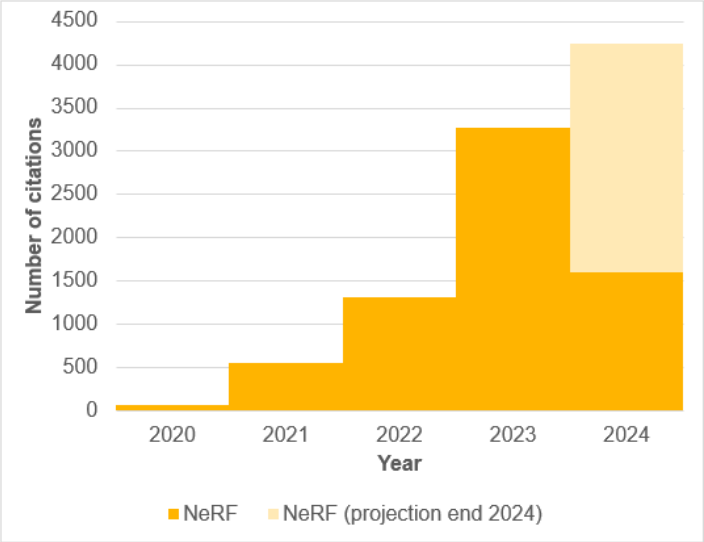

Figure 4 - **NeRF paper citations over the years.**

Following the introduction of NeRF, numerous radiance field architectures have been proposed. IBRNet [37] blends classical interpolation-based view synthesis with a non-scene-specific radiance field MLP. Optimized radiance field architectures been demonstrated to have real-time rendering capacities [38], [39], [40]. More efficient sampling strategies have been studied for faster rendering [41], anti-aliased and generalizable NeRF for boundless scenes [10], [42]. Many extensions of radiance fields have been proposed to extend the applications to other research fields. Large-scale NeRF extend the capacities of radiance fields to city-scale models [43], [44], [45]. Other contributions on scene understanding integrate NeRF into a scene graph [46], [47], for an editable volumetric representation. A large amount of NeRF studies are focused on more specific tasks such as avatar or face reconstruction [48], [49], [50].

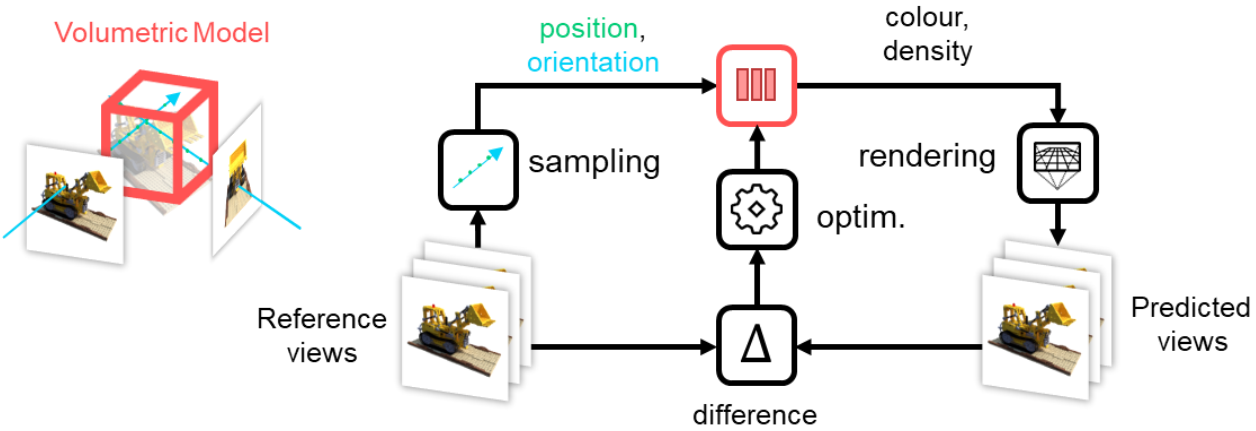

Figure 5 - Radiance Fields reconstruction pipeline.

Reconstruction of radiance fields from a limited set of views is possible, but failure may occur if the captured views are too sparse. The reconstruction process involves recovering a 5D function from 2D images, which is an under-resolved problem. Significant advances have been made in improving radiance field stability with regularizations, which are rules to constrain the radiance field function to converge towards a coherent model. Three main

regularizations have greatly improved radiance field stability. The homogeneity regularization, proposed as Total Variation (TA) by Lombardi et al. [51], encourages the model to have homogeneous zones. This means that the model must feature compact objects with diffuse colour. The regularization can be applied on 3D points [8], [52], [53] or encouraged on adjacent pixels of the rendered views [12], [54], [55]. Sparsity regularization encourages the emptiness of the model, thereby reducing the occurrence of unstructured artefacts. Beta-loss [8], [51], [51], Cauchy-loss [8], [39], [56] and entropy loss [53], [54] have been demonstrated to be efficient losses for sparsity regularization. Finally, appearance regularization encourages renders to appear correct, with the use of a trained CNN [12] or GAN [11].

	Regularization type				
	Homogeneity		Sparsity		Appearance
Method	sample	ray	sample	ray	image
Neural Volumes	TV		Beta		
NeRF					
NSVF			Beta		
Baking-NeRF				Cauchy	
PlenOctree				Cauchy	
MIP-NeRF 360	Dist				
DirectVOXGO	RGB			Entropy	
Plenoxels	TV		Beta	Cauchy	
RegNeRF		Depth			Colour
InfoNeRF		Gain		Entropy	
Dense Depth Priors		Depth			

Table 1 - Overview of regularisations for radiance fields

Spatially Encoded Radiance Fields

NeRF represents a significant advance in the field of computer vision, but it comes at a cost. The MLP, a key component of the NeRF model, is a relatively slow network to infer, requiring multiple inferences for a single pixel. One of the major advances in radiance field research has been the simplification of the implicit radiance function. To reduce the complexity load on the MLP, it can be spatially decomposed into smaller functions, as demonstrated in [57]. Yu et al. demonstrated that the radiance function can be reduced to a simple MLP-free representation, encoding density directly and orientation-dependent colour in a simple parametrization [8], [56].

Other research has investigated the use of a voxel grid to store feature vectors in the three-dimensional space. The feature vectors are trained alongside the MLP [53] and inputted to the MLP depending on the sampling location. The volumetric model information is divided into smaller batches that are more focused on local features. Consequently, equal or higher rendering quality to NeRF can be achieved with a smaller MLP architecture, resulting in faster rendering. As demonstrated in [39] and [58], feature vectors can be stored in sparse voxel grids. Müller et al. developed Instant-NeRF [59] a multi-resolution feature voxel grid-

based radiance field, which enables the rendering of higher-quality radiance with faster rendering speeds.

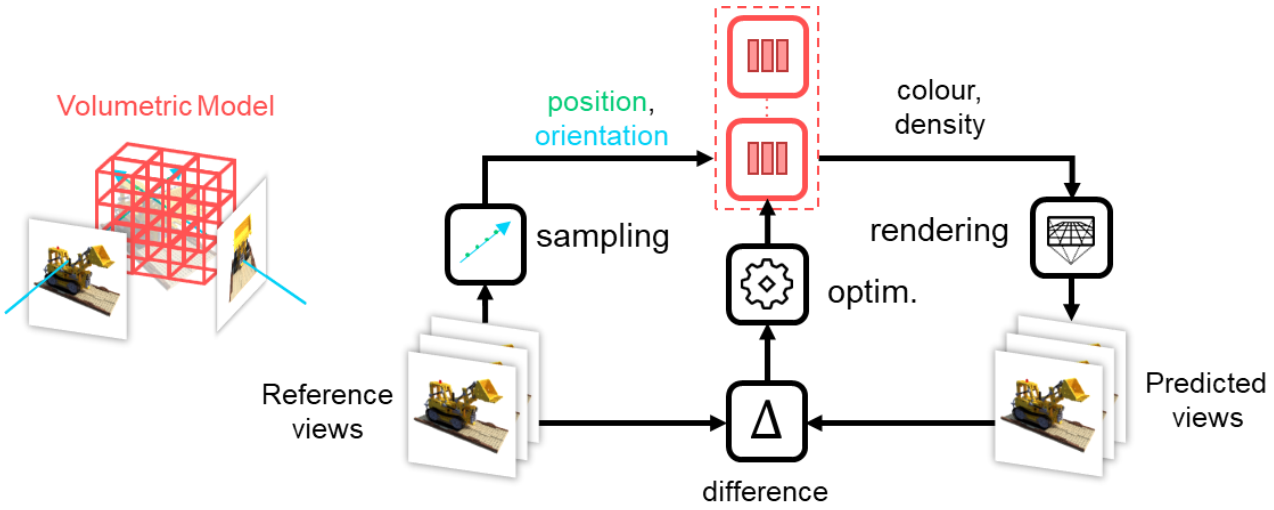

Figure 6 - Example representation of spatially encoded radiance field.

3D Gaussian Splatting (3DGS) is a method published in 2023 [13] that has rapidly gained recognition in volumetric reconstruction research. It is commonly associated with radiance field methods, and is a differentiable point cloud-based approach for learning-based volumetric rendering. The 3DGS model is composed of 3D Gaussians with geometry, orientation–dependant colour and density trainable parameters. Novel views are rendered through rasterization of the Gaussians onto the new view image plane, which is a faster process than ray casting, while maintaining most properties of radiance field rendering. In contrast, previous learning-based point cloud rendering approaches [32], 3DGS generates and prunes points during reconstruction, and is not dependent on a dense point cloud initialization.

Figure 7 - Illustration of the temporal flickering in Gaussian Splatting renderings of adjacent frames on the Carpark scene [60]

Radiance field methods are flexible yet powerful trainable representations of 3D scenes. Intuitively, training radiance field models using successive video frames as input results in a 4D representation of a dynamic scene. While this is true, the lack of temporal constraints associated with the underlying under-resolution of the radiance field reconstruction results in temporal artefacts during rendering of dynamic scenes. Figure 7 illustrates the flickering that can occur with sparse input views. The phenomenon of flickering is particularly evident in reconstructed zones with a higher degree of shape ambiguity, where numerous potential reconstructions could satisfy the reference images rendering. Consequently, windows reflections and occluded areas exhibit more pronounced flickering artefacts.

Dynamic Radiance Fields

The integration of the temporal dimension in a radiance field-based volumetric representation offers two benefits. Firstly, it increases temporal homogeneity, reinforcing spatial information with temporal redundancy. Secondly, it addresses temporary occluded areas. Deformation-based approaches achieve this by dividing the dynamic radiance field into a spatial radiance field and a dynamic deformation field [49], [61], [62]. All temporal instants respect morphologically consistent changes. Similar proposals have been published for dynamic 3DGS [63], [64]. Other dynamic radiance field papers demonstrate excellent performance by inputting the temporal dimension to a first MLP [15] or the feature grid [65].

In contrast to NeRF-based methods, 3DGS features a fully explicit volumetric model that can be more directly extended to the temporal dimension. Yang et al. [66] extend the 3D Gaussians with a temporal dimension and constrain them to coherent movement. In SpaceTime Gaussian Feature Splatting (STG) [67], polynomials are trained to represent the 3D Gaussians movement, forcing smooth movement.

METHODS COMPARISON

The comparison of Radiance Field methods is a challenging task, given it is an evolving field. Evaluation of methods is often conducted on older datasets that may not fully reflect the capabilities of state-of-the-art methods. Many methods have metadata prerequisites, such as the scene bounding box or depth maps. Moreover, a significant number of radiance field implementations are designed for specific content, such as full frontal views, inside captures and concentric views. We evaluate the methods on scenes of the MUVOD Dataset [68], a compilation of multiview video sequences from various sources. The sequences feature varied content, environments, and capture rigs, and include scenes focusing on human interaction.

The evaluation dataset comprises 14 sequences made of between 8 and 30 views. The view calibration is conducted using COLMAP [18]. One middle view is excluded from the training and retained for evaluation purposes for each sequence. Following training, the evaluation view is rendered and compared with the reference. The evaluation metrics employed are PSNR, SSIM and LPIPS. The higher the PSNR and SSIM the better, the lower the LPIPS the better. The reconstruction time is the training time of the volumetric representation. The rendering time is the average rendering time for a frame. The memory usage is the average size of the volumetric model for a sequence.

The following methods were evaluated: Plenoxels [8], Nerfacto [9], 3DGS [13] and STG [67]. Plenoxels and Nerfacto were selected to represent spatially encoded radiance fields. Nerfacto is a community-driven implementation of Instant-NeRF [59]. Plenoxels is trained on a high-resolution grid of 1024^3 voxels. Both Nerfacto and Plenoxels utilize the COLMAP calibration sparse point cloud, as described in [69], to initialize their bounding boxes. 3DGS is the trained for 7000 iterations. The STG method is the only dynamic method of the tested methods and is a dynamic extension of 3DGS. It is evaluated on five frames, and the results are averaged to be equivalent to a single frame, as with the other methods. Other state-of-the-art methods such as Mip-NeRF [10] or GANeRF [11] are not evaluated despite their high rendering quality due to their lengthy training times. The methods are trained with default parameters on a Nvidia A100 GPU, and the results are shown in Table 2.

Method	*PSNR ↑*	*SSIM ↑*	*LPIPS ↓*	*Recon time ↓*	*Memory Usage ↓*
Plenoxels	22.963	0.810	0.368	45 min	2,85 Go
Nerfacto	24,662	0,816	0,261	**8 min**	167 Mo
3DGS	27.803	0.866	0.220	**8 min**	140 Mo
STG	**29.523**	**0.916**	**0.177**	10 min	**5 Mo**

Table 2 - Methods comparison results. Average reconstruction and rendering metrics over 14 Multiview sequences.

Compared to all three other methods, Plenoxels has lower performance for each metric. Plenoxels' regular voxel grid stores feature vectors at the same resolution throughout the scene. This results in suboptimal information resolution for reconstructing foreground content, which has a higher resolution in reference images. In Figure 8, the foreground content is visibly blurred for the Plenoxels renders. This data structure also results in long training times and higher memory consumption compared to other methods.

Nerfacto performs significantly worse than 3DGS and STG on all rendering quality metrics, but has the lowest reconstruction time tied with Nerfacto and a comparable memory usage to 3DGS. Renders of two scenes are shown in Figure 8. The rendering quality of Nerfacto is very high for simple content like the car, but the reconstruction is particularly unstable for more complex scenes. For example, the people in the background in the lower image are poorly reconstructed due to occlusions from the people in the foreground.

3DGS and STG have the best rendering quality compared to the other two methods. The 3D Gaussians are a powerful and flexible representation. Even for difficult scenes, the training converges to a coherent geometry due to the point cloud initialization with COLMAP. 3DGS and related work are a promising solution for reliable high quality volumetric video with real-time rendering.

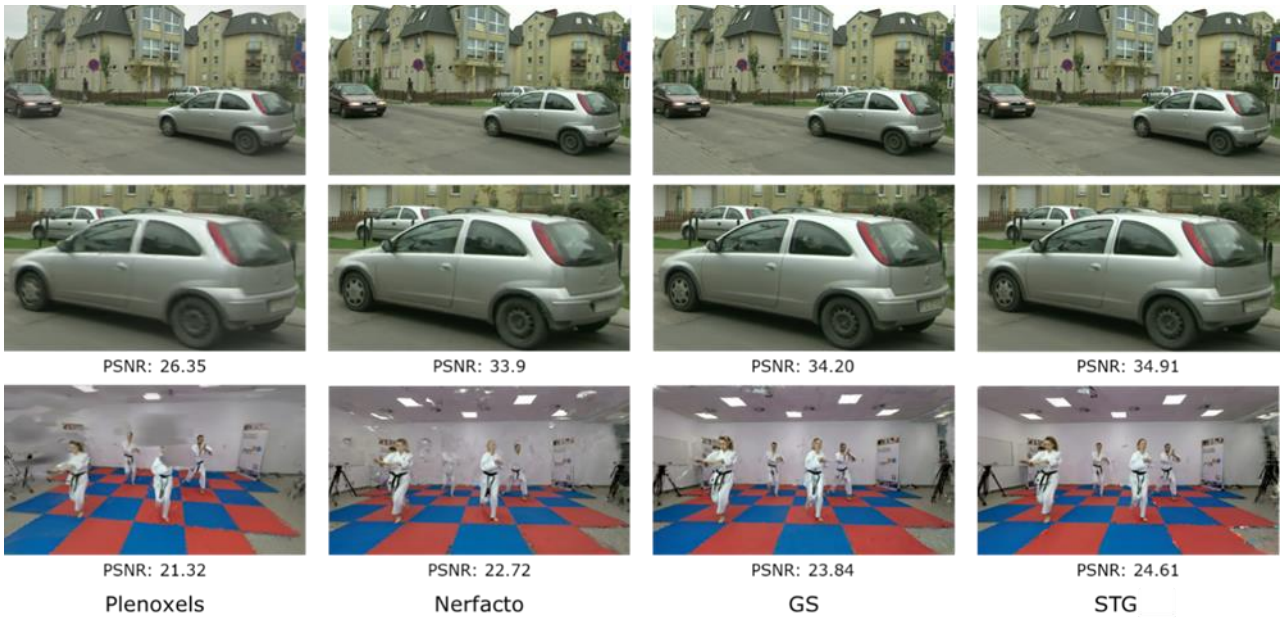

Figure 8 – Renders from sequences PoznanStreet [60] and MartialArts [70]

STG renders have better quality for every metric compared to all three other methods. This demonstrates the benefit of temporal information that can resolve occlusion ambiguities. If

part of the scene is visible in the other training frames, this information constrains the occluded area to coherent content. The average reconstruction time of STG for a single frame is comparable to 3DGS and Nerfacto, and could be reduced by training with more frames. STG has significantly lower memory consumption than the other methods evaluated. A single model is used for all five frames, resulting in an optimized, memory efficient model that could be further optimized by training on a larger number of frames.

The image metrics used evaluate images independently, and the video rendering of STG is not compared to the reference video. However, there is a strong gain in temporal stability of the renders compared to training independent frames with 3DGS. Figure 7 illustrates the rendering of independent frames and Figure 9 illustrates the simultaneous training of multiple frames with STG. Temporal flickering is visible for 3DGS in areas of ambiguous geometry, while renders are coherent for STGFS in static areas. While this can have an important impact on subjective quality, it is a missing piece of information for PSNR, SSIM, and LPIPS, which are the metrics classically used for volumetric video quality evaluation.

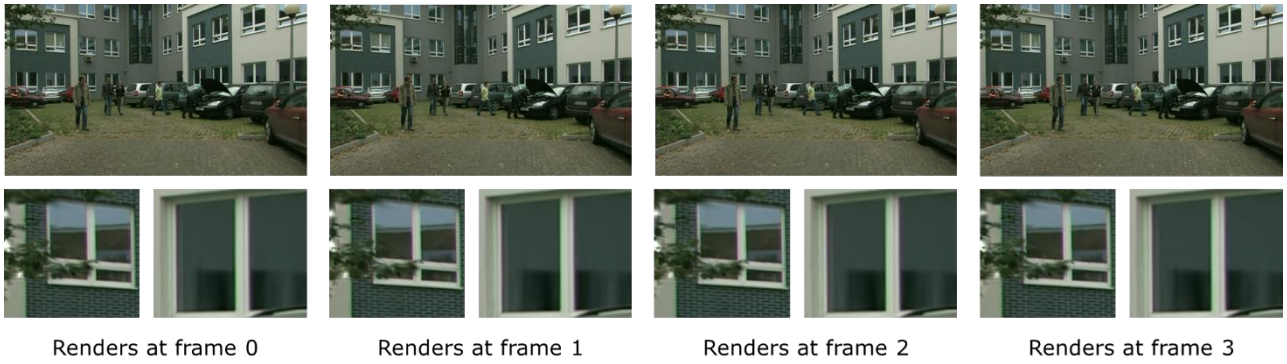

Figure 9 - Illustration of the temporal stability in Spacetime Gaussian Feature Splatting renderings of adjacent frames models on the Carpark scene [60]

CONCLUSIONS

In this paper, we provided an overview of volumetric video, focusing on recent advances in radiance fields for real-time free navigation of natural content. Advances in radiance field training and rendering optimization, reconstruction stability, and temporal expansion were detailed. The performance of state-of-the-art radiance field methods was evaluated in terms of objective metrics, training complexity, and memory usage on a multiview video dataset of complex scenes.

Since the introduction of NeRF in 2020, volumetric video has steadily evolved towards reliable application in real-world use cases. Radiance field techniques keep improving and are close to maturity for widespread use. Volumetric video could be the long-awaited answer to the lack of engaging content on immersive displays, helping content providers create immersive experiences with minimal production costs.

REFERENCES

[1] A. Smolic, « 3D video and free viewpoint video—From capture to display », *Pattern Recognition*, vol. 44, nº 9, p. 1958-1968, 2011, doi: https://doi.org/10.1016/j.patcog.2010.09.005.

[2] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, et A. Hilton, « Computer vision for sports: Current applications and research topics », *Computer Vision and Image Understanding*, vol. 159, p. 3-18, 2017, doi: https://doi.org/10.1016/j.cviu.2017.04.011.

[3] M. Hackett, B. Makled, E. Mizroch, S. Venshtain, et M. Mccoy-Thompson, « Volumetric Video and Mixed Reality for Healthcare Training », mai 2022.

[4] S. Orts-Escolano *et al.*, « Holoportation: Virtual 3d teleportation in real-time », in *Proceedings of the 29th annual symposium on user interface software and technology*, 2016, p. 741-754.

[5] « Intel True View - Intel in Sports ». [En ligne]. Disponible sur: https://www.intel.com/content/www/us/en/sports/technology/true-view.html

[6] C. Fehn, « Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV », *Proc SPIE*, vol. 5291, mai 2004.

[7] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, et R. Ng, « Nerf: Representing scenes as neural radiance fields for view synthesis », in *European Conference on Computer Vision*, 2020, p. 405-421.

[8] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, et A. Kanazawa, « Plenoxels: Radiance Fields Without Neural Networks », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, p. 5501-5510.

[9] M. Tancik *et al.*, « Nerfstudio: A Modular Framework for Neural Radiance Field Development », in *ACM SIGGRAPH 2023 Conference Proceedings*, in SIGGRAPH '23. 2023.

[10] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, et P. P. Srinivasan, « Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields », in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 5855-5864.

[11] B. Roessle, N. Müller, L. Porzi, S. R. Bulò, P. Kontschieder, et M. Nießner, « GANeRF: Leveraging Discriminators to Optimize Neural Radiance Fields », *ACM Trans. Graph.*, vol. 42, nº 6, nov. 2023, doi: 10.1145/3618402.

[12] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. M. Sajjadi, A. Geiger, et N. Radwan, « RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs », *CoRR*, vol. abs/2112.00724, 2021, [En ligne]. Disponible sur: https://arxiv.org/abs/2112.00724

[13] B. Kerbl, G. Kopanas, T. Leimkühler, et G. Drettakis, « 3D Gaussian Splatting for Real-Time Radiance Field Rendering », *ACM Transactions on Graphics*, vol. 42, nº 4, juill. 2023, [En ligne]. Disponible sur: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

[14] M. Ye, M. Danelljan, F. Yu, et L. Ke, « Gaussian grouping: Segment and edit anything in 3d scenes », *arXiv preprint arXiv:2312.00732*, 2023.

[15] L. Song *et al.*, « Nerfplayer: A streamable dynamic scene representation with decomposed neural radiance fields », *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, nº 5, p. 2732-2742, 2023.

[16] J. Zhang *et al.*, « Editable Free-viewpoint Video Using a Layered Neural Representation », *arXiv e-prints*, p. arXiv-2104, 2021.

[17] A. M. Barros, M. Michel, Y. Moline, G. Corre, et F. Carrel, « A Comprehensive Survey of Visual SLAM Algorithms », *Robotics*, vol. 11, nº 1, 2022, doi: 10.3390/robotics11010024.

[18] J. L. Sch"onberger et J.-M. Frahm, « Structure-from-Motion Revisited », in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[19] Z. Wang, S. Wu, W. Xie, M. Chen, et V. A. Prisacariu, « NeRF–: Neural Radiance Fields Without Known Camera Parameters », *arXiv e-prints*, p. arXiv:2102.07064, févr. 2021.

[20] Y. Jeong, S. Ahn, C. Choy, A. Anandkumar, M. Cho, et J. Park, « Self-calibrating neural radiance fields », in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 5846-5854.

[21] S.-F. Chng, S. Ramasinghe, J. Sherrah, et S. Lucey, « Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation », in *European Conference on Computer Vision*, 2022, p. 264-280.

[22] Z. Liu, P. An, S. Liu, et Z. Zhang, « Arbitrary view generation based on DIBR », in *2007 International Symposium on Intelligent Signal Processing and Communication Systems*, 2007, p. 168-171.

[23] M. Levoy et P. Hanrahan, « Light field rendering », in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, p. 441-452.

[24] S. J. Gortler, R. Grzeszczuk, R. Szeliski, et M. F. Cohen, « The lumigraph », in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, p. 453-464.

[25] A. Krizhevsky, I. Sutskever, et G. E. Hinton, « Imagenet classification with deep convolutional neural networks », *Advances in neural information processing systems*, vol. 25, 2012.

[26] J. F. N. R. L. Z. N. H. J. Maraval, « MUSE: A Multi-view Synthesis Enhancer », *To be published in EUSIPCO*, 2023.

[27] P. Hedman, J. Philip, T. Price, J.-M. Frahm, G. Drettakis, et G. Brostow, « Deep blending for free-viewpoint image-based rendering », *ACM Transactions on Graphics (TOG)*, vol. 37, nº 6, p. 1-15, 2018.

[28] W. Liu, Z. Piao, J. Min, W. Luo, L. Ma, et S. Gao, « Liquid Warping GAN: A Unified Framework for Human Motion Imitation, Appearance Transfer and Novel View Synthesis », in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, oct. 2019.

[29] R. Huang, S. Zhang, T. Li, et R. He, « Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis », in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, oct. 2017.

[30] G. Riegler et V. Koltun, « Free view synthesis », in *European Conference on Computer Vision*, 2020, p. 623-640.

[31] G. Riegler et V. Koltun, « Stable view synthesis », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 12216-12225.

[32] D. Rückert, L. Franke, et M. Stamminger, « Adop: Approximate differentiable one-pixel point rendering », *ACM Transactions on Graphics (TOG)*, vol. 41, nº 4, p. 1-14, 2022.

[33] B. Mildenhall *et al.*, « Local light field fusion: Practical view synthesis with prescriptive sampling guidelines », *ACM Transactions on Graphics (TOG)*, vol. 38, nº 4, p. 1-14, 2019.

[34] J. Flynn *et al.*, « Deepview: View synthesis with learned gradient descent », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, p. 2367-2376.

[35] S. Wizadwongsa, P. Phongthawee, J. Yenphraphai, et S. Suwajanakorn, « Nex: Real-time view synthesis with neural basis expansion », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 8534-8543.

[36] M. Broxton *et al.*, « Immersive light field video with a layered mesh representation », *ACM Transactions on Graphics (TOG)*, vol. 39, nº 4, p. 81-86, 2020.

[37] Q. Wang *et al.*, « Ibrnet: Learning multi-view image-based rendering », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 4690-4699.

[38] S. J. Garbin, M. Kowalski, M. Johnson, J. Shotton, et J. Valentin, « Fastnerf: High-fidelity neural rendering at 200fps », in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 14346-14355.

[39] P. Hedman, P. P. Srinivasan, B. Mildenhall, J. T. Barron, et P. Debevec, « Baking neural radiance fields for real-time view synthesis », in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 5875-5884.

[40] B. Deng, J. T. Barron, et P. P. Srinivasan, « JaxNeRF: an efficient JAX implementation of NeRF, 2020 ».

[41] T. Neff *et al.*, « DONeRF: Towards Real-Time Rendering of Neural Radiance Fields using Depth Oracle Networks ». 2021.

[42] K. Zhang, G. Riegler, N. Snavely, et V. Koltun, « NeRF++: Analyzing and Improving Neural Radiance Fields ». 2020.

[43] M. Tancik *et al.*, « Block-NeRF: Scalable Large Scene Neural View Synthesis ». 2022.

[44] H. Turki, D. Ramanan, et M. Satyanarayanan, « Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, p. 12922-12931.

[45] M. Zhenxing et D. Xu, « Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields », in *The Eleventh International Conference on Learning Representations*, 2022.

[46] J. Ost, F. Mannan, N. Thuerey, J. Knodt, et F. Heide, « Neural scene graphs for dynamic scenes », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 2856-2865.

[47] M. Niemeyer et A. Geiger, « Giraffe: Representing scenes as compositional generative neural feature fields », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 11453-11464.

[48] M. Mihajlovic, A. Bansal, M. Zollhoefer, S. Tang, et S. Saito, « KeypointNeRF: Generalizing Image-based Volumetric Avatars using Relative Spatial Encoding of Keypoints ». arXiv, 2022. doi: 10.48550/ARXIV.2205.04992.

[49] K. Park *et al.*, « HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields ». 2021.

[50] A. Grigorev *et al.*, « StylePeople: A Generative Model of Fullbody Human Avatars ». 2021.

[51] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, et Y. Sheikh, « Neural volumes: Learning dynamic renderable volumes from images », *arXiv preprint arXiv:1906.07751*, 2019.

[52] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, et P. Hedman, « Mip-nerf 360: Unbounded anti-aliased neural radiance fields », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, p. 5470-5479.

[53] C. Sun, M. Sun, et H.-T. Chen, « Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, p. 5459-5469.

[54] M. Kim, S. Seo, et B. Han, « InfoNeRF: Ray Entropy Minimization for Few-Shot Neural Volume Rendering », déc. 2021.

[55] B. Roessle, J. T. Barron, B. Mildenhall, P. P. Srinivasan, et M. Nießner, « Dense depth priors for neural radiance fields from sparse input views », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, p. 12892-12901.

[56] A. Yu, R. Li, M. Tancik, H. Li, R. Ng, et A. Kanazawa, « Plenoctrees for real-time rendering of neural radiance fields », in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, p. 5752-5761.

[57] D. Rebain, W. Jiang, S. Yazdani, K. Li, K. M. Yi, et A. Tagliasacchi, « DeRF: Decomposed Radiance Fields », *arXiv preprint arXiv:2011.12490*, 2020.

[58] L. Liu, J. Gu, K. Z. Lin, T.-S. Chua, et C. Theobalt, « Neural sparse voxel fields », *Advances in Neural Information Processing Systems*, vol. 33, p. 15651-15663, 2020.

[59] T. Müller, A. Evans, C. Schied, et A. Keller, « Instant Neural Graphics Primitives with a Multiresolution Hash Encoding », *ACM Trans. Graph.*, vol. 41, n° 4, p. 102:1-102:15, juill. 2022, doi: 10.1145/3528223.3530127.

[60] D. Mieloch, A. Dziembowski, et M. Domański, « [MPEG-I Visual] Natural Outdoor Test Sequences », *Natural Outdoor Test Sequences, Brussels, Belgium*, 2020.

[61] A. Pumarola, E. Corona, G. Pons-Moll, et F. Moreno-Noguer, « D-nerf: Neural radiance fields for dynamic scenes », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, p. 10318-10327.

[62] J.-W. Liu *et al.*, « Devrf: Fast deformable voxel radiance fields for dynamic scenes », *Advances in Neural Information Processing Systems*, vol. 35, p. 36762-36775, 2022.

[63] G. Wu *et al.*, « 4d gaussian splatting for real-time dynamic scene rendering », *arXiv preprint arXiv:2310.08528*, 2023.

[64] Y. Liang *et al.*, « GauFRe: Gaussian Deformation Fields for Real-time Dynamic Novel View Synthesis », *arXiv preprint arXiv:2312.11458*, 2023.

[65] S. Fridovich-Keil, G. Meanti, F. R. Warburg, B. Recht, et A. Kanazawa, « K-planes: Explicit radiance fields in space, time, and appearance », in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, p. 12479-12488.

[66] Z. Yang, H. Yang, Z. Pan, X. Zhu, et L. Zhang, « Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting », *arXiv preprint arXiv:2310.10642*, 2023.

[67] Z. Li, Z. Chen, Z. Li, et Y. Xu, « Spacetime gaussian feature splatting for real-time dynamic view synthesis », *arXiv preprint arXiv:2312.16812*, 2023.

[68] B. Wei, J. Maraval, M. Outtas, K. Kpalma, N. Ramin, et L. Zhang, « MUVOD: Multi-view Video Object Segmentation Dataset ». Submission in progress, 2023. [En ligne]. Disponible sur: https://volumetric-repository.labs.b-com.com/#/muvod

[69] L. Z. M. J. N. RAMIN, « K3BO: Keypoint-based bounding box optimization for radiance field reconstruction from multi-view images », *ICME workshop on Immersive Media Compression*, 2023.

[70] D. Mieloch *et al.*, « [MIV] New natural content - MartialArts ». janvier 2023.