Bringing Reality to Virtual Worlds: An Evaluation of 3D Gaussian Splatting for Accessible 3D Reconstruction in Virtual Production

Chrysoula Kaloumenou
chryssa.kalou@gmail.com
Aalborg University
Copenhagen, Denmark

Figure 1: 3DGS model of AAU Copenhagen, Jawset Postshot

ABSTRACT

Virtual production (VP) relies on efficient and accessible 3D reconstruction techniques to seamlessly integrate real-world elements into digital scenes. This research investigates 3D Gaussian Splatting (3DGS) as a potential alternative to traditional photogrammetry, with a focus on utilizing readily available tools and mid-range capture devices. A variety of props and locations were scanned using a smartphone and a drone, and the resulting data was processed using both a local 3DGS implementation (Jawset Postshot) and a cloud-based platform (Luma AI). The generated 3D models were evaluated based on geometric accuracy, texture fidelity, and real-time rendering performance. Preliminary results indicate that both Postshot and Luma AI offer promising alternatives to photogrammetry, with varying strengths and weaknesses depending on the complexity of the scanned asset and capture method. While Luma AI excels in speed and ease of use, Postshot provides greater flexibility and customization options. This study highlights the potential of 3DGS as an accessible and efficient tool for 3D reconstruction in VP, and suggests future work exploring the capabilities of Neural Radiance Fields (NeRFs) for novel view synthesis in this context.

KEYWORDS

3D Gaussian Splatting, NeRF, Radiance Fields, Photogrammetry, Virtual Production, Scanning, Filmmaking

1 INTRODUCTION

Filmmaking is undergoing a profound transformation, thanks to the emergence of Virtual Production (VP) – a revolutionary approach that seamlessly merges physical and digital elements in real-time ([1], [2]). This paradigm shift is fueled by advancements in real-time rendering engines and LED walls, enabling filmmakers to craft immersive and visually stunning narratives by blurring the lines between tangible sets and computer-generated environments [2].

One of the critical challenges in VP lies in the seamless integration of real-world objects (props) into digital scenes, ensuring a cohesive and realistic visual experience [16]. Traditionally, this integration has relied heavily on 3D reconstruction techniques like photogrammetry, where 3D models are created from 2D images ([17], [18]). However, photogrammetry often struggles to accurately capture intricate geometries, reflective surfaces, and scenes with dynamic lighting conditions, which can introduce artifacts, inaccuracies, and time-consuming manual corrections [10].

This research explores two cutting-edge techniques that offer promising alternatives to photogrammetry: 3D Gaussian Splatting (3DGS) and Neural Radiance Fields (NeRFs). 3DGS, an evolution of traditional image processing, represents scenes using strategically positioned 3D Gaussian functions (splats) ([6], [8]). This approach offers a computationally efficient alternative, potentially enabling real-time or near-real-time reconstruction and rendering, making it particularly appealing for the fast-paced nature of virtual production [3]. NeRFs, on the other hand, leverage deep learning to represent scenes as continuous volumetric functions, allowing for the generation of novel views and perspectives not explicitly captured during scanning ([12], [14]). This capability, known as novel view synthesis, opens up exciting possibilities for virtual camera movements and creative exploration within the digital environment.

By investigating the performance of 3DGS and NeRFs in capturing and representing real-world objects and environments using

Figure 2: VP set, Image courtesy of ViZARTS

readily available tools and mid-range capture devices, this study aims to assess their suitability for a wider range of users and applications within the film industry. The research will delve into the geometric accuracy, texture fidelity, and real-time rendering performance of these techniques, comparing them to traditional photogrammetry and evaluating their potential to democratize 3D reconstruction in virtual production.

2 BACKGROUND

To understand the transformative impact of virtual production, it's essential to examine its rapid rise in the filmmaking industry, the pivotal role of digital twins, and the challenges in 3D reconstruction that this technology presents.

2.1 The Rise of Virtual Production in Filmmaking

Virtual Production (VP) is rapidly reshaping the filmmaking landscape, offering an innovative approach that seamlessly integrates physical and digital elements in real-time ([1], [2]). This paradigm shift is driven by advancements in real-time rendering engines and LED walls, which enable filmmakers to create immersive and visually stunning narratives by merging tangible sets with computer-generated environments [2].

Traditionally, this integration relied heavily on chroma keying (green screen technology), where actors performed in front of a green backdrop later replaced with digital backgrounds in post-production [20]. While green screens offered flexibility and cost-effectiveness, they lacked real-time feedback, limited interaction with virtual environments, and struggled with accurate lighting and reflections [16].

VP addresses these limitations by directly displaying virtual environments on set, fostering natural interaction between actors and their surroundings and providing instant visual feedback for filmmakers [2]. This real-time interaction streamlines the filmmaking process, promotes creative decision-making on the fly, and enhances the overall production experience. However, seamlessly integrating real-world objects and environments into these virtual scenes remains a challenge that hinges on the accuracy and efficiency of 3D reconstruction techniques.

2.2 The Role of Digital Twins in Virtual Production

In the world of VP, the concept of digital twins has emerged as a cornerstone for achieving realism and efficiency. A digital twin is a virtual replica of a physical object or environment, meticulously crafted to mirror its real-world counterpart in both form and function [7]. While digital twins have found applications in diverse industries, their significance in VP is particularly pronounced, especially during the production phase.

During production, digital twins enhance realism and interaction by allowing actors to engage with virtual objects as if they were physically present [20]. This not only enriches the actors' performances but also streamlines the integration of real and virtual elements, resulting in a more cohesive and believable final image.

Furthermore, digital twins facilitate real-time tracking and compositing, ensuring that virtual elements seamlessly align with physical objects and actors on set [1]. This real-time synchronization is essential for achieving convincing visual effects and minimizing the need for extensive post-production corrections.

The use of digital twins also enables virtual cinematography, empowering filmmakers to explore a wider range of camera angles and perspectives [2]. By utilizing virtual cameras within the digital twin environment, directors and cinematographers can experiment with shots that would be physically impossible or impractical to achieve on a traditional set. This not only expands creative possibilities but also optimizes production timelines and budgets.

2.3 Emerging Alternatives: NeRFs and 3D Gaussian Splatting

To overcome the limitations of photogrammetry, researchers are actively exploring emerging alternatives like Neural Radiance Fields (NeRFs) and 3D Gaussian Splatting (3DGS). NeRFs, pioneered by [14], employ deep learning to represent scenes as continuous volumetric functions. This approach has demonstrated exceptional capabilities in reconstructing intricate scenes with realistic lighting and reflections ([12], [23]). Notably, their ability to generate novel views not explicitly captured during scanning (novel view synthesis) has transformative potential for virtual camera movements and creative exploration in VP [19].

3DGS, rooted in traditional image processing techniques, represents scenes using strategically placed 3D Gaussian functions (splats) [6], [8]. This method offers a computationally efficient alternative, potentially enabling real-time or near-real-time reconstruction and rendering, a critical advantage for the fast-paced nature of VP [3]. Recent advancements in 3DGS have demonstrated its capacity to capture and render dynamic scenes effectively, further expanding its potential applications in VP.

While both NeRFs and 3DGS show immense promise, their practical implementation and usability within VP workflows, especially for users with varying levels of technical expertise and resources, remain an area ripe for exploration. Research into the performance, accessibility, and integration of these techniques in real-world VP scenarios is crucial to assess their potential to democratize 3D reconstruction and make it more accessible to a broader spectrum of filmmakers.

Table 1: Summary of capture devices, methods, and lighting conditions

Asset	Capture Device	Method	Lighting Conditions
Cornelius (gnome)	Samsung S22 Ultra, iPad	Photo, Video, Polycam	Ring light, Natural
Philodendron (plant)	Samsung S22 Ultra, iPad	Photo, Video, Polycam	Ring light, Natural
Contemplation room	Samsung S22 Ultra, iPad	Photo, Video, Polycam	Ring light, Natural
Bicycle	Samsung S22 Ultra, iPad	Photo, Video, Polycam	Daylight
AAU Facade	DJI Ryze Tello Drone	Video	Daylight

3 RESEARCH OBJECTIVES

This research aims to evaluate the potential of 3D Gaussian Splatting (3DGS) and Neural Radiance Fields (NeRFs) as viable alternatives to traditional photogrammetry for 3D reconstruction in Virtual Production (VP) workflows. The study will prioritize the exploration of 3DGS, leveraging readily available and accessible tools, with the goal of democratizing this technique for a wider range of users and applications.

The research will address the following key objectives:

- **Assess** the geometric accuracy and texture fidelity of 3D models generated using 3DGS, utilizing free or online software and mid-range capture devices (e.g., smartphones) [6].
- **Document** the technical challenges and limitations encountered during the processing of datasets using accessible tools and workflows, aligning with the findings of [3] regarding computational resource constraints.
- **Compare** the processing time and resource requirements of 3DGS and NeRFs to traditional photogrammetry, evaluating the ease of integrating 3DGS models into existing VP pipelines, as explored by [15] and [21].
- **Evaluate** the real-time rendering performance of the generated models within a virtual production environment (e.g., Unreal Engine), building upon the work of [16] in assessing real-time visualization capabilities.

By addressing these objectives, this research will provide valuable insights into the potential of these new technologies for democratizing 3D reconstruction in virtual production, making it more accessible and affordable for a wider range of filmmakers and content creators.

4 METHODOLOGY

This research investigates the potential of 3D Gaussian Splatting (3DGS) and NeRFs in VP workflows using accessible tools and mid-range capture devices. The study focuses on evaluating the quality, efficiency, and suitability of the generated models for integration into VP pipelines. Starting from 3DGS processing, the research is intending to then delve into NeRFs processing, and compare the results each of these approaches yielded.

4.1 Rationale for Prioritising 3DGS over NeRFs

While NeRFs predate 3DGS and have shown remarkable potential in 3D reconstruction, this research chose to prioritize the exploration of 3DGS for two primary reasons. First, the computational demands of NeRFs, particularly the training of neural networks, can lead to significantly longer processing times compared to 3DGS. Given the time constraints of this study, starting with the faster 3DGS approach allowed for a more comprehensive exploration of its capabilities within the available timeframe.

Second, the development of 3DGS from traditional image processing techniques, specifically the concept of Gaussians, presents an intriguing academic curiosity. Investigating how these established methods can compete with cutting-edge AI and ML approaches like NeRFs offers valuable insights into the evolving landscape of 3D reconstruction. By prioritizing 3DGS, this research aims to shed light on the potential of both traditional and emerging techniques in the context of virtual production.

4.2 Data Acquisition

To evaluate the performance of 3DGS and NeRFs under varying conditions and levels of complexity, a diverse set of assets were selected for scanning, progressing in order of increasing complexity. This allowed for an iterative learning process, where insights and techniques learned from scanning simpler assets could be applied to more challenging ones.

The scanning process began with a simple object: a ceramic Christmas gnome (Cornelius) chosen for its solid form, matte surface, and intricate details. Next, a Philodendron plant was scanned to explore the challenges of capturing organic forms and fine details, as well as to address the practical issue of limited availability of 3D plant assets for virtual production. The complexity was then further increased by scanning the indoor contemplation room at AAU Copenhagen, a complex scene with diverse objects and varying textures.

Subsequently, a bicycle was scanned due to its highly reflective surfaces and intricate geometry, known to be challenging for 3D reconstruction techniques [16]. Finally, the facade of the AAU building in Copenhagen was scanned using a drone, pushing the limits of the technique by attempting to reconstruct a large-scale outdoor environment with reflective glass surfaces using limited resources.

For image and video capture of the props and the contemplation room, a Samsung S22 Ultra smartphone was utilized. All props and the contemplation room were also scanned using an iPad Pro (12.9", 4th generation) equipped with LiDAR technology and the Polycam app (v3.3.1) [22]. Aerial footage of the AAU building facade was acquired using a DJI Ryze Tello drone.

4.3 Data Processing

Each captured dataset was processed using two pipelines:

(1) **Luma AI:** Photos and videos were uploaded to Luma AI's online platform, which utilizes proprietary algorithms to generate 3D models. Luma AI is a widely recognized and

user-friendly platform in the 3D reconstruction community, offering a streamlined process for creating 3DGS models. Its cloud-based approach allows for quick and convenient processing, making it a popular choice for users seeking rapid results.

(2) **Jawset Postshot:** Local processing on a dedicated workstation using Jawset Postshot software to generate 3D models through 3D Gaussian Splatting (3DGS) [9]. Project settings in Postshot were optimized to achieve a balance of quality and processing time (usually under 24 hours). Specific parameters adjusted included the number of images used, features per image, and training steps. Postshot, on the other hand, caters to users who prioritize data security and customization. Its local processing ensures that sensitive data remains on the user's machine, addressing potential concerns in virtual production environments. Additionally, Postshot is the first freely available software that simplifies the implementation of 3DGS, offering a more accessible alternative to the complex process outlined in the original research that introduced the concept of 3D Gaussian Splatting [9]. It provides a range of adjustable parameters, allowing for fine-tuning of model quality and greater control over the reconstruction process.

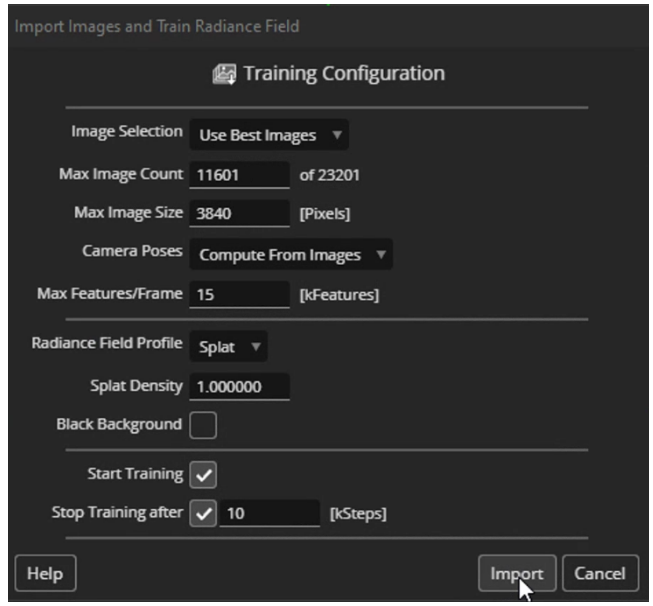

Figure 3: The import options of Jawset Postshot

4.4 Hardware Specifications

The dedicated workstation used for local processing comprised:

- **Operating System:** Windows 10 Pro
- **Processor:** AMD Ryzen 7 3800XT 8-Core Processor (3.89 GHz)
- **Installed Memory (RAM):** 32.0 GB (2 x 16 GB G.Skill Intl DDR4 SDRAM @ 2133 MHz)
- **GPU:** NVIDIA GeForce RTX 3090

4.5 Evaluation and Analysis

The resulting 3D models were evaluated based on qualitative and quantitative assessments, encompassing the following criteria:

- **Geometric Accuracy:** Visual inspection and comparison to reference images, with potential use of quantitative metrics (e.g., mean distance to reference).
- **Texture Fidelity:** Visual inspection and comparison to reference images, considering level of detail, sharpness, and realism of material representation.
- **Real-Time Rendering Performance:** Preliminary tests in Unreal Engine 5.3 using the Luma AI plugin for Gaussian Splatting, assessing frame rates and visual quality. The Luma AI plugin was chosen due to its compatibility with 3DGS models and its ability to leverage the computational efficiency of Gaussian Splatting for real-time rendering.
- **Ease of Integration:** Assessment of the process of incorporating 3DGS models into Unreal Engine 5.3 using the Luma AI plugin.

For each criterion, a comparative analysis was conducted between the Postshot and Luma AI models for each asset, considering the capture method and resolution.

5 RESULTS

This section presents the results of the 3D model reconstruction and evaluation process using a local 3DGS implementation (Jawset Postshot), a cloud-based 3DGS platform (Luma AI), and a mobile app (Polycam). While the initial research plan included an exploration of Neural Radiance Fields (NeRFs), this aspect has been deferred to future work due to time constraints.

A total of 160 3D reconstructions were attempted across various assets, capture methods, and settings. The analysis that follows focuses on geometric accuracy, texture fidelity, real-time rendering performance, and ease of integration. The two primary platforms used for 3DGS-based 3D reconstruction, Postshot and Luma AI, offer distinct approaches to the process. Postshot provides users with a range of adjustable parameters (*Figure 3*), including the choice between using all captured images or a subset of the "best" images selected automatically by the software. Users can also control the maximum number of images used (when using the "Best Images" option), the maximum number of features per image, and the number of training steps (iterations of the algorithm refining the model). This flexibility allows for experimentation and potential optimization of model quality. In contrast, Luma AI's platform is fully automated, requiring no user input or parameter adjustments beyond uploading the captured data.

5.1 Cornelius the Gnome

5.1.1 ***Luma AI.*** Luma AI consistently produced high-quality models of Cornelius with good geometric accuracy and texture fidelity. Processing time was consistently fast (30-40 minutes), regardless of capture settings. The best results were obtained from high-resolution photos (108MP) and videos (FHD 30fps). However, Luma

Table 2: Comparison of Luma AI and Postshot for 3D Reconstruction in Virtual Production

Feature	Luma AI	Postshot	Polycam
Processing Time	Consistently fast (30-40 minutes)	Varies widely (minutes to 10+ hours)	Very fast (2-3 minutes)
Ease of Use	Very High (extremely user-friendly, minimal steps, accessible on mobile devices)	High (simple interface, but requires experimentation to understand parameter impact)	High (user-friendly mobile app)
Customization	Limited (no adjustable parameters)	High (offers control over image selection, features, training steps)	Limited (no adjustable parameters)
Model Quality	Generally good, occasional artifacts	Potential for higher quality with optimization, but also more prone to errors	Varies widely (based on the size and type of surface scanned)
Handling of High-Resolution Data	Limited (struggles with 8K video)	Limited (crashes with 108MP photos and some high-res videos)	Limited by LiDAR capabilities
Hardware Requirements	None (cloud-based processing)	Dependent on model complexity and desired quality (can be demanding on resources)	Requires LiDAR-equipped device
Ideal Use Case	Time-sensitive projects, users seeking quick results, users with limited computational resources	Projects prioritizing high quality and detail, users willing to experiment for optimal results, users with powerful hardware	Quick capture of objects with simple geometry, suitable for mobile use

Figure 4: Cornelius: Photo & 3DGS model, Jawset Postshot

AI's models occasionally exhibited minor artifacts and inconsistencies in color reproduction, particularly when using lower resolution video with AI enhancement.

5.1.2 ***Jawset Postshot.*** Postshot consistently outperformed Luma AI in terms of geometric accuracy and texture fidelity for Cornelius. The best results were obtained from high-resolution photos (108MP) and videos (FHD 30fps), capturing intricate details like the scratches on the hat. However, processing times varied significantly, ranging from 1 hour to over 8 hours depending on the chosen settings. Attempts to process videos in FHD 60fps and UHD 30fps resulted in software crashes.

High-resolution photos (108MP) processed in Postshot, using 81 images and 15 features per image, resulted in a model with a processing time ranging from 1 hour (10k training steps) to >8 hours (100k training steps). This model exhibited high quality but had some minor distortions. Photos with lower resolution processed in Postshot, using 20-100 features per image, resulted in models with processing times ranging from 21 minutes (10k training steps) to >22 hours (500k training steps). These models were of high quality but still demonstrated some floaters and artifacts.

Videos in FHD 30fps processed in Postshot, using 2200 images and 15 features per image, resulted in models with processing times ranging from just over 1 hour (10k training steps) to >5 hours (500k training steps). These models were of very high quality, with reflections and textures well-represented. Videos in HD 30fps processed in Postshot, using all 2200 images and 15 features per image, resulted in a model with a processing time of more than 12 hours (10k training steps). This model was of very high quality, with few flaws.

5.1.3 ***Polycam.*** Two LiDAR scans were performed on Cornelius using Polycam, one using 112 frames and the other using 92 frames. Both models were exceptionally accurate in terms of geometric fidelity; however, the 92-frame one was more photorealistic due to its higher texture resolution.

5.2 Philodendron Plant

5.2.1 ***Luma AI.*** Luma AI produced good results for the Philodendron plant, with the best model achieved using UHD 60fps video capture. However, some artifacts and inconsistencies in texture representation were observed, particularly around the leaf edges.

Figure 5: Philodendron plant: 3DGS model, Jawset Postshot

5.2.2 ***Jawset Postshot***. Postshot's models of the Philodendron plant showed superior geometric accuracy compared to Luma AI, especially in capturing the subtle curvature of the leaves. However, Postshot also struggled with artifacts and inconsistencies in texture representation, particularly around the leaf edges. The 108MP photogrammetry model exhibited high quality, but processing failures occurred when attempting to import the full-resolution images.

5.2.3 ***Polycam***. Two LiDAR scans were performed on the Philodendron plant using Polycam, one with natural lighting and the other one using a ring light. The model with natural lighting was slightly better, with fewer artifacts and shadows. However, none of the models managed to create an accurate reconstruction.

5.3 Contemplation Room

5.3.1 ***Luma AI***. Luma AI faced difficulties in accurately representing the entire Contemplation Room, with issues such as floaters, misplaced objects, and distortions being common. The 8K 24fps video scan resulted in a particularly poor model. However, Luma AI performed better in capturing individual elements within the room, such as the slim plant and the moss wall with the log, when scanned separately.

5.3.2 ***Jawset Postshot***. Postshot struggled with the complexity of the Contemplation Room, often resulting in incomplete or distorted reconstructions, especially with higher resolution videos. Processing the 8K video led to repeated failures.

Postshot also captured individual elements within the room, with varying degrees of success. The moss-and-log scans were generally of higher quality than the scans of the banana plant, slim plant, and chair.

5.3.3 ***Polycam***. The LiDAR scan of the entire Contemplation Room using Polycam resulted in a model where everything felt fluid and distorted, far from a photorealistic representation. The

Figure 6: Contemplation room: 3DGS model, Luma AI

Figure 7: Bike: 3DGS model, Luma AI

chair had disappeared, leaving only its shadow and a hint of the legs. Individual scans of the chair were also not photorealistic and exhibited holes and distortions.

5.4 Bicycle

5.4.1 ***Luma AI***. Luma AI's models of the bicycle generally exhibited better geometric accuracy and texture fidelity compared to Postshot, particularly in capturing the reflective surfaces and intricate details of the bike. The best results were achieved using high-resolution photos (108MP) and videos (FHD 60fps). Luma AI also produced acceptable results with lower resolution videos (FHD 30fps and HD 30fps), although some artifacts were present around the wheels and steering bar.

5.4.2 ***Jawset Postshot***. Postshot struggled with the highly reflective surfaces of the bicycle, often producing models with artifacts and distortions. However, Postshot's models showed some improvement with higher resolution images and increased processing steps, indicating the potential for better results with further optimization. Notably, attempts to process the 108MP photos of the bike in Postshot repeatedly crashed the software, highlighting potential limitations in handling large datasets or complex scenes.

The default resolution photos processed in Postshot, using 15-100 features per image, resulted in models with processing times ranging from 31 minutes (10k training steps) to over 18 hours (300k training steps). These models exhibited varying degrees of quality, with some being potentially usable if placed in the background, while others suffered from floaters, artifacts, and an oil-painting-like appearance.

5.4.3 ***Polycam***. The LiDAR scan of the bike using Polycam resulted in a low-quality model with the wall being scanned better than the bike itself. This suggests that Polycam might not be the most suitable tool for capturing objects with complex geometries and reflective surfaces.

5.5 AAU Building Facade

5.5.1 ***Luma AI***. Luma AI produced surprisingly good results for the AAU building facade, considering the challenges posed by its large scale and reflective glass surfaces. Initially, processing each of the two drone video datasets individually did not yield satisfactory results. However, when the datasets from both flights were combined, the resulting model exhibited good overall quality and clarity. The lettering on the facade was clear, and the reflections were well-represented.

5.5.2 ***Jawset Postshot***. Similar to Luma AI, the two separate datasets captured by the drone failed to yield quality results when processed individually in Postshot. However, when combined, the resulting model captured the overall structure and some details of the facade, but suffered from more noise and artifacts compared to Luma AI's model. The limited resolution of the drone footage (720p) likely contributed to these limitations.

6 DISCUSSION

The preceding sections have detailed the methodology and results of the experiments conducted to evaluate the potential of 3D Gaussian Splatting (3DGS) as an alternative to photogrammetry in virtual production. This section will delve deeper into the comparative analysis of the two primary 3DGS pipelines used (Luma AI and Jawset Postshot), interpreting the findings, discussing their implications for virtual production workflows, and identifying limitations and avenues for future research.

6.1 Comparative Analysis and Interpretation

The results presented in Section 5 reveal both the potential and limitations of Jawset Postshot and Luma AI for 3D reconstruction in virtual production. Both techniques demonstrated varying degrees of success in capturing geometric accuracy and texture fidelity, depending on the complexity of the scanned asset and the capture settings.

Simpler Objects (Cornelius the Gnome): For simpler objects, both Postshot and Luma AI produced high-quality models. However, Postshot consistently performed better, particularly when utilizing high-resolution photos or videos. This suggests that Postshot, with its ability to leverage more input data and user-defined parameters, might be more adept at capturing intricate details and subtle variations in surface geometry. Luma AI, while generally producing good results, occasionally exhibited minor artifacts and inconsistencies, especially with lower-resolution inputs. This could indicate that its fully automated approach might not always be optimal for capturing the nuances of complex textures or lighting conditions.

Complex Organic Objects (Philodendron Plant): The Philodendron plant, with its complex leaf structures, posed a greater challenge for both techniques. While Postshot maintained its advantage in geometric accuracy, both platforms struggled with artifacts and inconsistencies in texture representation, particularly around leaf edges. This suggests that capturing fine details and organic forms remains a challenge for 3D reconstruction techniques, regardless of the underlying algorithm. Notably, the failure of Postshot to process the full-resolution 108MP photogrammetry model indicates potential limitations in handling large datasets.

Complex Scenes (Contemplation Room): The Contemplation Room, with its diverse objects and varying lighting conditions, proved to be the most challenging asset for both techniques. Luma AI struggled with the overall scene reconstruction, producing a particularly poor model from the 8K video. However, it performed better when capturing individual elements within the room, highlighting its potential suitability for reconstructing isolated objects rather than complex scenes. Postshot also faced difficulties with the complexity of the scene, often resulting in incomplete or distorted reconstructions, especially with higher resolution videos. This suggests that both Postshot and Luma AI might require further optimization or specialized techniques to effectively handle highly complex scenes.

Highly Reflective Objects (Bicycle): The bicycle, with its highly reflective surfaces and intricate geometry, presented a significant challenge for both techniques. Luma AI's models generally outperformed Postshot in capturing the reflections and fine details, particularly when using high-resolution photos and videos. Postshot struggled with reflections, often producing artifacts and distortions. However, Postshot showed some improvement with higher resolution images and increased processing steps, suggesting that further optimization could yield better results. The repeated crashes encountered when processing 108MP photos in Postshot highlight potential limitations in handling large datasets or complex scenes with this software.

Large-Scale Environments (AAU Building Facade): The AAU building facade, captured using a drone, presented unique challenges due to its large scale and reflective glass surfaces. Both Luma AI and Postshot produced usable results when combining footage from multiple drone flights. Luma AI's model exhibited slightly better overall quality and clarity, while Postshot's model suffered from more noise and artifacts. This suggests that Luma AI's cloud-based processing might be more adept at handling large datasets and complex scenes, while Postshot might require more powerful hardware or further optimization for such scenarios.

In summary, the comparative analysis reveals that both Luma AI and Postshot offer viable alternatives to traditional photogrammetry for 3D reconstruction in virtual production, each with its own strengths and weaknesses depending on the specific use case. Luma AI's speed and ease of use make it a compelling option for time-sensitive projects and users with limited resources, while Postshot's flexibility and customization potential cater to those prioritizing high quality and detail. The choice between the two platforms ultimately depends on the specific requirements and constraints of the virtual production workflow. Further exploration of these techniques, including their real-time rendering performance, will shed light on their broader applicability and potential to democratize 3D reconstruction in the film industry.

6.2 Real-Time Rendering Performance

Preliminary tests in Unreal Engine 5.3 using the Luma AI plugin for Gaussian Splatting, chosen for its compatibility with 3DGS models and efficient real-time rendering capabilities (Luma AI Documentation), indicate that both Postshot and Luma AI models generally perform well in terms of real-time rendering. The average frame rates for both Luma AI and Postshot models were consistently above 30fps, meeting the typical requirements for real-time rendering in virtual production. While these results suggest that while 3DGS models can achieve real-time rendering performance, optimization may be necessary for more complex or high-resolution assets.

During initial testing, an issue with flickering 3DGS assets was encountered when importing them into Unreal Engine. This occurred only on the dedicated virtual production workstation (equipped with rear projection and VIVE trackers) and not on other machines. Troubleshooting steps, including reinstalling software and drivers, did not resolve the issue. However, it was discovered that importing the model into the tracked scene in Unreal Engine eliminated the flickering. The exact cause of this issue remains unclear, but the workaround allowed for the successful use of 3DGS models in the VP environment.

6.3 Ease of Integration and Processing Time

Both 3DGS and Luma AI models were seamlessly integrated into Unreal Engine 5.3 using the Luma AI plugin for Gaussian Splatting. Postshot models, exported in the PLY format, were directly compatible with the plugin, requiring no additional format conversion or material adjustments. The integration process involved importing the PLY file into Unreal Engine, assigning materials, and adjusting lighting and rendering settings as needed.

6.4 Implications for Virtual Production

The findings of this research have several implications for the adoption and utilization of 3DGS in virtual production workflows. The speed and ease of use of Luma AI's online platform make it a compelling option for time-sensitive projects or situations where computational resources are limited. Its ability to consistently produce usable results within a short timeframe, regardless of the complexity of the asset, is a significant advantage in fast-paced production environments. However, the lack of customization options may limit its appeal for users who require greater control over the reconstruction process or seek to achieve the highest possible quality for specific assets.

Conversely, Postshot's 3DGS implementation, while more time-consuming and computationally demanding, offers greater flexibility and control through its adjustable parameters. This allows for fine-tuning of the reconstruction process to prioritize specific aspects of model quality, such as geometric accuracy or texture fidelity. The ability to experiment with different settings and optimize for specific needs could be invaluable for projects where achieving the highest possible quality is paramount, even at the expense of longer processing times.

The choice between Luma AI and Postshot, therefore, depends on the specific requirements and constraints of the VP workflow. For projects with tight deadlines or limited resources, Luma AI's speed and simplicity might be the deciding factor. However, for projects that prioritize customization and have the resources to invest in longer processing times, Postshot's 3DGS implementation could be a more suitable choice.

The results also highlight the importance of selecting appropriate capture methods and settings for optimal results. While video capture generally yielded better texture fidelity than photogrammetry for both techniques, the optimal resolution and frame rate varied depending on the specific asset and the desired level of detail. High-resolution images and videos, while potentially leading to better geometric accuracy, also increased processing times and, in some cases, caused software instability. This suggests that users need to carefully consider the trade-offs between quality, efficiency, and the limitations of their hardware and software when choosing capture settings.

Furthermore, the challenges encountered in capturing complex scenes like the contemplation room underscore the need for further research and development in 3DGS techniques. While both techniques showed promise in capturing individual elements within the room, their performance in reconstructing the entire scene was less satisfactory. This suggests that capturing complex scenes with diverse objects and lighting conditions remains a challenge for 3DGS, and further advancements are needed to improve its robustness and accuracy in such scenarios.

6.5 Limitations and Future Work

This research provides valuable insights into the potential of 3DGS for 3D reconstruction in virtual production, but it is not without limitations. The scope of this study was constrained by the limited number and types of assets scanned, which may not fully represent the diversity of objects and environments encountered in real-world virtual production scenarios. The selection of assets primarily focused on static objects and indoor environments, with limited exploration of dynamic scenes or large-scale outdoor environments. Additionally, the use of a single smartphone for capture and a single dedicated workstation for processing could introduce biases and limit the generalizability of the findings. Future research could address these limitations by expanding the dataset to include a wider variety of assets, such as dynamic objects, humans in motion, and diverse outdoor environments. Exploring the performance of 3DGS with different capture devices, including professional-grade cameras and specialized 3D scanners, could also provide valuable insights into the scalability and adaptability of this technique.

Furthermore, while this study focused on 3DGS, future research should revisit the investigation of Neural Radiance Fields (NeRFs) for novel view synthesis in virtual production, as originally planned. The ability of NeRFs to generate photorealistic novel views from a limited set of input images could be a game-changer for virtual production, enabling greater flexibility and creative control in scene composition and camera movement.

Additionally, exploring the potential of other cutting-edge techniques could open up new avenues for 3D reconstruction in VP. For instance, Dynamic Gaussian Splatting, which aims to improve the representation of dynamic scenes ([5], [13]), could address the limitations of 3DGS in capturing objects in motion. SMERFs (Sparse Multiscale Encoding of Radiance Fields), which have shown promising results in reconstructing high-quality 3D scenes from sparse input views [4], could be particularly relevant for virtual production workflows where capturing dense image sets might be impractical or time-consuming. The concept of Vast Gaussians, introduced by [11], offers a potential solution for reconstructing large-scale scenes, as demonstrated in the AAU building facade experiment, where combining multiple datasets was necessary to achieve satisfactory results.

By continuing to investigate and refine these cutting-edge methods, we can unlock new possibilities for 3D reconstruction and further enhance the creative potential of virtual production.

6.6 Potential for Democratization of 3D Reconstruction

The findings of this research suggest that 3DGS, particularly through accessible tools like Luma AI and Postshot, has the potential to democratize 3D reconstruction in virtual production by making it more accessible to users with limited resources and technical expertise. The use of readily available tools and mid-range capture devices, coupled with the relatively fast processing times and ease of integration, could empower a wider range of filmmakers and content creators to incorporate high-quality 3D models into their virtual productions.

However, further research and development are needed to address the limitations of 3DGS, particularly in handling complex scenes and high-resolution data. Additionally, the development of more user-friendly interfaces and streamlined workflows could further enhance the accessibility and usability of these tools for non-experts.

7 CONCLUSION

This research successfully explored the potential of 3D Gaussian Splatting (3DGS) and Luma AI as accessible alternatives to traditional photogrammetry for 3D reconstruction in virtual production. The results demonstrate that both techniques offer promising results, with 3DGS (implemented in Postshot) often achieving higher geometric accuracy and texture fidelity for simpler objects, while Luma AI excels in speed and ease of use, making it ideal for time-sensitive projects. Both techniques faced challenges with complex objects and environments, highlighting the need for further research and development in this area.

The findings of this research suggest that tools like Jawset Postshot and Luma AI have the potential to democratize 3D reconstruction in virtual production by making it more accessible and affordable for a wider range of filmmakers and content creators. The use of readily available tools and mid-range capture devices, coupled with the relatively fast processing times and ease of integration, could empower a wider range of individuals and studios to incorporate high-quality 3D models into their virtual productions, fostering greater creativity and innovation in the field.

Future work will focus on evaluating the potential of Neural Radiance Fields (NeRFs) for novel view synthesis in virtual production, as well as exploring other emerging techniques like SMERFs. By continuing to investigate and refine these cutting-edge methods, we can unlock new possibilities for 3D reconstruction and further enhance the creative potential of virtual production.

In conclusion, this research contributes valuable insights into the evolving landscape of 3D reconstruction in virtual production, paving the way for a more accessible, efficient, and creative future for filmmakers.

REFERENCES

[1] M. Bremer and B. Groß. 2023. Virtual Production: A Holistic Approach to Filmmaking. Springer.

[2] R. Carvalho, M. Oliveira, and A. Coelho. 2022. *Virtual Production for Film and Television: A Complete Guide.* CRC Press.

[3] K. Chaitanya, S. Melam, K. N. Ravi, and D. Rajagopal. 2022. Real-Time 3D Reconstruction and Rendering using Gaussian Splatting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.* 315–324.

[4] D. Duckworth, B. Trevithick, and J. T. Barron. 2023. SMERF: Streamable Memory Efficient Radiance Fields for Real-Time Large-Scene Exploration. *arXiv preprint arXiv:2312.07541v2* (2023).

[5] C. Guanjun, S. Melam, K. N. Ravi, and D. Rajagopal. 2023. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. *arXiv preprint arXiv:2310.08528v2* (2023).

[6] P. Hedman, E. Ringaby, and O. Olsson. 2021. Efficient 3D Gaussian Splatting for Real-Time Rendering. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games.* 1–11.

[7] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks. 2020. Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology* 29 (2020), 36–52.

[8] A. Kerbl, P. Hedman, K. Takayama, and O. Olsson. 2023. PlenOctrees for Real-Time Rendering of Neural Radiance Fields. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–15.

[9] A. Kerbl, V. Kopanas, M. Leimkŏ0fchler, and G. Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics (TOG)* 41, 4 (2023), 1–17.

[10] J. Lin, Y. Wu, C. Zheng, and Y. Liu. 2024. Challenges and Opportunities in 3D Reconstruction for Virtual Production. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–15.

[11] J. Lin, Y. Wu, C. Zheng, and Y. Liu. 2024. VastGaussian: Vast 3D Gaussians for Large Scene Reconstructions. *arXiv preprint arXiv:2402.17427v1* (2024).

[12] S. Liu, X. Gu, B. Shi, S. B. Kang, and J. Yu. 2020. Neural Sparse Voxel Fields. In *Advances in Neural Information Processing Systems (NeurIPS).*

[13] J. Luiten, F. Yu, T. Li, P. Hedman, K. Takayama, R. Yang, J. T. Barron, M. J. Black, and A. Kanazawa. 2023. Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis. *arXiv preprint arXiv:2308.09713v1* (2023).

[14] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision.* Springer, 405–421.

[15] J. Park, U. Sinha, J. T. Barron, S. Bouaziz, D. B. Goldman, S. M. Seitz, and R. Martin-Brualla. 2022. DeRF: Decomposed Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 8857–8866.

[16] J. Reimão, M. M. Oliveira, and H. Araújo. 2023. Virtual Production: A Systematic Review of the State of the Art. *Computers & Graphics* 108 (2023), 100123.

[17] F. Remondino and S. El-Hakim. 2006. Image-based 3D modelling: A review. *The Photogrammetric Record* 21, 115 (2006), 269–291.

[18] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 1. IEEE, 519–528.

[19] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein. 2022. Implicit Neural Representations with Periodic Activation Functions. In *Advances in Neural Information Processing Systems*, Vol. 34. 7462–7473.

[20] B. Smith. 2018. *Virtual Production: The Future of Filmmaking.* Routledge.

[21] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, and R. Ng. 2022. Block-NeRF: Scalable Large Scene Neural View Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 8867–8876.

[22] Q. Xu, X. Wang, C. Li, Y. Chen, and L. Chen. 2020. A LiDAR and Photogrammetry Fusion Method for 3D Reconstruction. *Remote Sensing* 12, 10 (2020), 1607.

[23] Y. Zhang, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and R. Ng. 2022. NeRF++: Analyzing and Improving Neural Radiance Fields. In *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition. 8847–856.

ACKNOWLEDGMENTS

Henrik, your guidance, support, and endless patience made this project possible. Jesper & Ivan, thank you for all the help. K, M, P and G, here's to you.