Still2Scene: Hybrid Gaussian Environments for Virtual Production

Xiaohan Sun[1] and Carol O'Sullivan[1]

[1]Trinity College Dublin, Ireland

Abstract

Virtual production often requires rapidly generated background environments that support real-time rendering and limited camera motion. While volumetric 3D Gaussian splatting provides high visual fidelity, it is computationally expensive, whereas planar billboard representations are efficient but lack geometric depth. We present a hybrid Gaussian scene representation that converts a single image into a lightweight navigable environment for virtual production. The proposed approach combines volumetric foreground Gaussians with planar Gaussian primitives for distant regions, while additional billboard assets can be synthesized using diffusion-based image prompting. The resulting hybrid scene can be deployed directly in Unreal® Engine for real-time rendering. This workflow enables fast environment generation from images and provides a practical middle-ground between static backplates and fully authored 3D environments in virtual production pipelines.

CCS Concepts

• ***Computing methodologies*** → *Image-based rendering; Rendering; Scene modeling;*

1. Introduction

Constructing digital environments for virtual production typically requires significant manual effort, including asset modeling, scene assembly, and environment optimization. While modern real-time engines such as Unreal® Engine enable high-quality rendering, producing production-ready environments remains time-consuming. Approaches are therefore needed that can rapidly generate navigable environments directly from visual inputs such as a single image. Recent Gaussian-based scene representations are a promising solution for efficient real-time rendering. However, while volumetric 3D Gaussian splatting provides high visual fidelity and parallax, it often requires a large number of primitives, thereby increasing computational cost [YDH*25, MDL*25]. In contrast, planar or billboard-based Gaussian representations are more efficient but provide limited geometric depth under camera motion [HYC*24, WB24].

We propose a hybrid Gaussian representation that converts a single image into a lightweight navigable environment for virtual production. Our approach reconstructs foreground structures using volumetric 3D Gaussians while representing distant regions with planar 2D Gaussian primitives. Diffusion-based image prompting is further used to generate additional billboard assets that enrich the environment. The resulting scene can be deployed directly in Unreal Engine, enabling rapid environment creation for virtual production tasks such as pick-up shots or quick background iteration. Our contributions are as follows:

- A hybrid Gaussian scene representation that combines volumetric foreground Gaussians with planar Gaussian primitives to enable lightweight navigable environments derived from a single image.
- A workflow to convert images into virtual production environments by integrating layered Gaussian reconstruction and diffusion-generated billboard assets in Unreal Engine.

2. Method

We propose a hybrid Gaussian scene generation pipeline that converts a single image into a navigable environment. The key idea is to assign different Gaussian representations to spatial layers of the scene in order to balance geometric fidelity and rendering efficiency. The pipeline consists of three stages: image preprocessing and scene decomposition, hybrid Gaussian asset generation, and scene composition. Figure 1 illustrates the overall workflow.

Given an input image I, we first enhance image quality using Real-ESRGAN [WYW*18], producing an improved image $I' = f_{sr}(I)$. A monocular depth estimator then predicts a depth map $D = f_{depth}(I')$, which is used to separate the image into foreground and background regions based on relative depth values. Foreground regions are reconstructed using volumetric Gaussian primitives following recent single-image 3D Gaussian reconstruction approaches [MDL*25]. Distant regions are approximated using planar Gaussian primitives generated with Image-GS [ZLK*25].

To enrich distant scene structures, additional billboard assets are synthesized using diffusion-based image prompting based on IP-Adapter [YZL*23]. The background image I_b serves as the visual prompt together with a text description T. Generated images are converted into billboard textures and represented as planar Gaussian primitives. All assets are assembled into a hybrid scene representation: $\mathcal{S} = G_{3D}(I_f) \cup G_{2D}(I_b) \cup G_{2D}(B(I_b, T))$. The resulting scene is imported into Unreal Engine and deployed as a navigable environment within a virtual production workflow.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.2312/egp.20261012

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

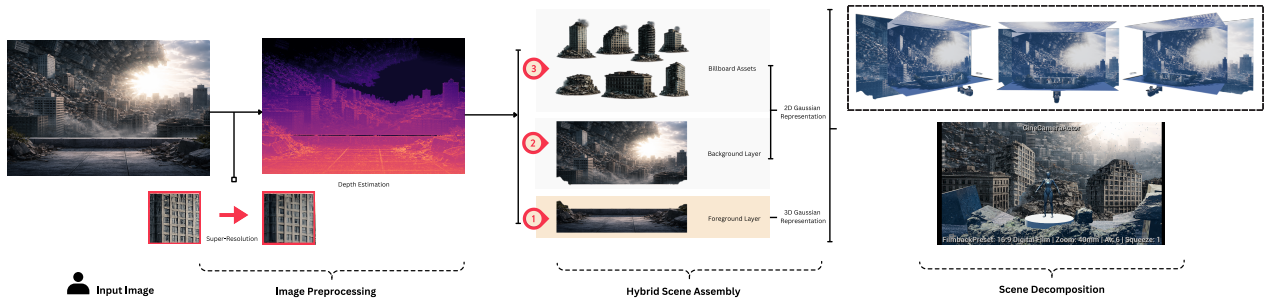

Figure 1: *Pipeline overview. A single input image is enhanced and decomposed into foreground and background layers using depth estimation. Foreground structures are reconstructed with volumetric 3D Gaussians, while distant regions and generated billboards are represented using planar 2D Gaussian primitives before being assembled into a hybrid scene for real-time rendering.*

Method	# Gaussians	FPS
3D Gaussian (Full)	2.36M	45.6
2D Gaussian (Planar)	0.10M	113.4
Hybrid + Billboards	1.35M	83.8

Table 1: *Rendering performance comparison between different scene representations. Runtime statistics are averaged over approximately 20 seconds of interactive camera navigation in Unreal Engine 5.5 using an NVIDIA RTX 4090 Laptop GPU.*

3. Evaluation

We compare three scene configurations: (1) a full 3D Gaussian reconstruction, (2) a purely planar 2D Gaussian representation, and (3) our hybrid representation combining volumetric foreground Gaussians with planar background primitives and diffusion-generated billboard assets.

Table 1 reports rendering performance measured in Unreal Engine using an nDisplay virtual production setup. All statistics were collected during interactive camera navigation and averaged over approximately 20 seconds of runtime.

The full 3D Gaussian reconstruction contains 2.36M primitives and achieves 45.6 FPS, providing strong geometric parallax but incurring high computational cost. The purely planar 2D Gaussian representation contains 0.10M primitives and achieves 113.4 FPS, offering high rendering efficiency but limited depth cues during camera motion.

Our hybrid scene combines 1.18M foreground volumetric Gaussians with 0.10M planar background primitives and seven diffusion-generated billboard assets, resulting in 1.35M primitives and 83.8 FPS. This configuration reduces the number of volumetric primitives while maintaining real-time rendering performance suitable for virtual production environments generated from a single image. Additional visual results and video demonstrations are provided in the supplemental materials.

4. Conclusions

We introduced a hybrid Gaussian workflow that converts a single image into a lightweight navigable environment for virtual production. By combining volumetric foreground Gaussians with planar background primitives and generated billboard assets, the approach provides a practical middle-ground between static image backplates and fully authored 3D environments.

Experiments demonstrate that the hybrid representation supports real-time rendering while reducing volumetric complexity. Future work will investigate perceptual evaluation through user studies with virtual production practitioners to assess visual realism and usability during camera operation.

Acknowledgement

This work was conducted with the financial support of the Research Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[HYC*24] HUANG B., YU Z., CHEN A., GEIGER A., GAO S.: 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers* (2024), pp. 1–11.

[MDL*25] MESCHEDER L., DONG W., LI S., BAI X., SANTOS M., HU P., LECOUAT B., ZHEN M., DELAUNOY A., FANG T., ET AL.: Sharp monocular view synthesis in less than a second. *arXiv preprint arXiv:2512.10685* (2025).

[WB24] WEISS S., BRADLEY D.: Gaussian billboards: Expressive 2d gaussian splatting with textures. *arXiv preprint arXiv:2412.12734* (2024).

[WYW*18] WANG X., YU K., WU S., GU J., LIU Y., DONG C., QIAO Y., CHANGE LOY C.: Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (2018), pp. 0–0.

[YDH*25] YU H.-X., DUAN H., HERRMANN C., FREEMAN W. T., WU J.: Wonderworld: Interactive 3d scene generation from a single image. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 5916–5926.

[YZL*23] YE H., ZHANG J., LIU S., HAN X., YANG W.: Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721* (2023).

[ZLK*25] ZHANG Y., LI B., KUZNETSOV A., JINDAL A., DIOLATZIS S., CHEN K., SOCHENOV A., KAPLANYAN A., SUN Q.: Image-gs: Content-adaptive image representation via 2d gaussians. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers* (2025), pp. 1–11.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.