Supplementary Material to: EnvMap-GS: Two-Stage Outdoor Gaussian Reconstruction with Background-to-Environment Map Baking

Deborah Pintani, Ariel Caputo, Noah Lewis, Marc Stamminger, Fabio Pellacini, Andrea Giachetti

1 Introduction

In this document, we present additional material for the paper "EnvMap-GS: Two-Stage Outdoor Gaussian Reconstruction with Background-to-Environment Map Baking".

In detail, we include:

- **Comparative evaluation:** A quantitative and qualitative comparison of our method against an hybrid neural/GS technique, like ScaffoldGS.
- **Ablation study:** The complete data on an analysis of the threshold R_i used to segment the foreground from the background. Moreover, we show an example image of the background initialization strategy, using geodesic sampling.
- **Environment map analysis:** A detailed quantitative evaluation of the environment maps generated from our background Gaussians, comparing different projection methods (Cylindrical vs. Spherical) and Spherical Harmonics bands.

2 Comparison of novel view synthesis with Scaffold GS

For quality comparison, we also considered the light fields estimated with ScaffoldGS (ScafGS) [4], which requires a different renderer for neural Gaussians, therefore is not usable in the standard GS viewers and VR applications. Despite this fact, considered that it provides very good novel view synthesis performances, we have compared the test images generated by this method with our results on the selected benchmark. The average results are reported in Table 1. Quality metrics are quite similar. Figure 1 shows that our method provides better results in the sky region (green and yellow boxes), while ScaffoldGS appear more accurate for near objects (red boxes).

3 Ablation study on thresholds

Table 2 reports the novel view synthesis metrics across different foreground/background threshold values (R_i). The highlighted columns indicate the thresholds selected by our automatic heuristic. As shown, our heuristic identifies values where the quality metrics are either optimal or near-optimal.

We observe that the system is generally robust: for most scenes, the performance remains stable across a range of thresholds. Lower thresholds (e.g., $10m$) tend to perform worse as they force foreground Gaussians to model too much of the background, while very high thresholds can lead to less efficient memory usage without significantly improving the final novel view synthesis. The thresholds chosen by our method consistently land in this robust range, demonstrating that the heuristic is a reliable proxy for the scene's geometric structure.

In Figure 2 we visualize the initialization of background points. We use a geodesic pattern to populate the spherical shell defined by R_i and R_o, ensuring uniform coverage for distant background elements. Points beyond R_o are projected onto the outer sphere to represent the sky, while points within the shell are sampled to account for mid-distance structures.

4 Quantitative evaluation results using Environment Maps

Table 3 reports the standard novel view synthesis metrics obtained when replacing the background Gaussians with the generated environment maps, that we include for completeness. As anticipated in the main text, these metrics are noticeably lower than those achieved with the full 3D Gaussian representation.

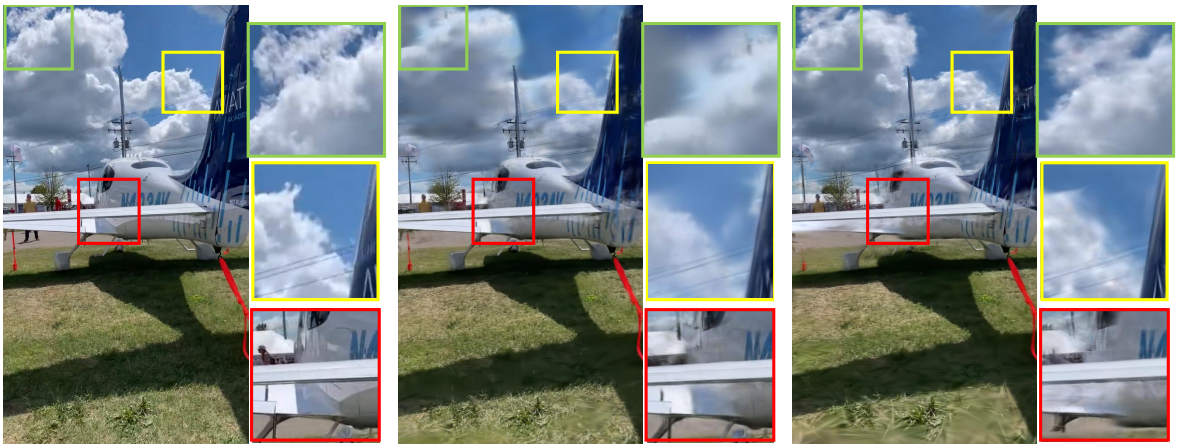
This degradation is expected: standard metrics evaluate pixel-perfect alignment against ground-truth images captured from different camera translations. Since a static environment map lacks depth, it cannot reproduce motion parallax, leading to geometric misalignments that severely penalize metrics.

In the table, we compare two map generation methods: our proposed Cylindrical Slit-Scan approach (which is then warped to equirectangular) and a direct Spherical rendering approach, adapted from [3] [1] [2]. We ultimately adopted the Cylindrical method in our pipeline because the direct Spherical approach introduces severe polar distortions that degrade visual quality.

Furthermore, we evaluated the impact of Spherical Harmonics (SH) bands. When baking Gaussians into a static texture, higher SH bands (which encode view-

	Person				Fields				Plane				Francis				Average			
	SSIM	PSNR	LPIPS	N°	SSIM	PSNR	LPIPS	N°	SSIM	PSNR	LPIPS	N°	SSIM	PSNR	LPIPS	N°	SSIM	PSNR	LPIPS	N°
Ours	**0,831**	**29,22**	**0,230**	1.7M	**0,760**	**25,58**	**0,231**	2.8M	0,726	21,66	0,311	1.7M	0,898	**29,20**	0,231	0.8M	0,804	**26,41**	0,251	1.8M
ScaGS	0,822	27,89	0,246	0,5M	0,757	25,52	0,237	0.6M	**0,743**	**21,92**	**0,298**	0.6M	**0,908**	28,66	**0,207**	0.2M	**0,808**	26.00	**0,247**	0.5M

Table 1: Comparison of novel views generated with our method and ScaffoldGS.



GT ScaffoldGS (LPIPS: 0.332) OURS (LPIPS: 0.350)

Figure 1: Comparison of rendered images with our method and ScaffoldGS.

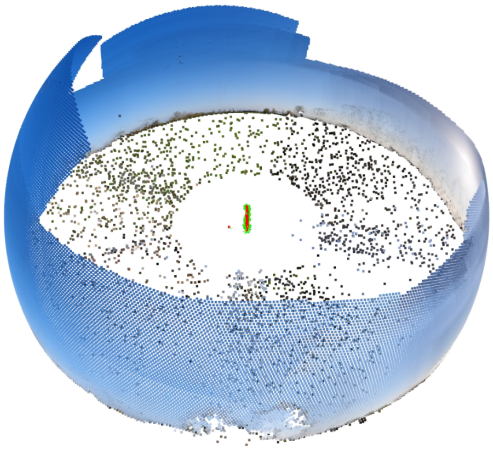


Figure 2: Visualization of the background initialization. Points corresponding to infinity (e.g., the sky) are projected onto the outer sphere in a geodesic pattern, while other distant elements are sampled within the spherical shell.

dependent reflections) can introduce baked-in artifacts. Our tests confirm that limiting the evaluation to the base diffuse color (SH0) or the first band (SH1) generally provides the most stable and artifact-free environment maps.

References

[1] Bai, J., Huang, L., Guo, J., Gong, W., Li, Y., Guo, Y.: 360-gs: Layout-guided panoramic gaussian splatting for indoor roaming. arXiv preprint arXiv:2402.00763 (2024)

[2] Huang, L., Bai, J., Guo, J., Li, Y., Guo, Y.: On the error analysis of 3d gaussian splatting and an optimal projection strategy. arXiv preprint arXiv:2402.00752 (2024)

[3] inuex35: 360 gaussian splatting (2026). URL `https://github.com/inuex35/360-gaussian-splatting`

[4] Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20,654–20,664 (2024)

Person								
SSIM	0,824	0,828	0,831	0,831	0,830	0,829	0,826	0,821
PSNR	29,027	29,145	29,216	29,212	29,153	29,110	28,964	28,605
LPIPS	0,240	0,233	0,230	0,230	0,230	0,231	0,233	0,238
n. (Mill.)	1,67	1,72	1,71	1,68	1,66	1,62	1,56	1,51
BG/FG threshold	10	20	30	40	50	60	70	80
Fields								
SSIM	0,743	0,755	0,756	0,758	0,759	0,759	0,761	0,760
PSNR	24,697	25,338	25,438	25,465	25,515	25,504	25,538	25,575
LPIPS	0,255	0,240	0,240	0,236	0,234	0,234	0,232	0,231
n. (Mill.)	2,56	2,78	2,88	2,84	2,85	2,83	2,87	2,86
BG/FG threshold	10	20	30	40	50	60	70	80
Plane								
SSIM	0,690	0,723	0,726	0,726	0,721	0,725	0,726	0,726
PSNR	20,091	21,483	21,592	21,657	21,500	21,571	21,593	21,635
LPIPS	0,360	0,316	0,312	0,311	0,315	0,312	0,310	0,313
n. (Mill.)	1,40	1,71	1,76	1,74	1,71	1,73	1,72	1,73
BG/FG threshold	10	20	30	40	50	60	70	80
Francis								
SSIM	0,876	0,898	0,897	0,897	0,892	0,888	0,887	0,882
PSNR	24,614	29,198	29,003	29,003	28,678	28,388	28,091	27,402
LPIPS	0,258	0,231	0,232	0,232	0,237	0,241	0,241	0,245
n. (Mill.)	0,70	0,81	0,82	0,79	0,76	0,71	0,71	0,69
BG/FG threshold	10	20	30	40	50	60	70	80

Table 2: Ablation study on different distances, in meters, used to separate foreground and background in our method. Distances corresponding to infinity are always set to 200. The best threshold value is highlighted.

		Cylindical + Equirectangular				Spherical				
		SH 0	SH 1	SH 2	SH 3	SH 0	SH 1	SH 2	SH 3	Threshold
Person	SSIM	0,606	0,605	0,603	0,601	0,443	0,443	0,443	0,443	30
	PSNR	22,20	22,21	22,21	22,12	9,87	9,89	9,93	10,00	
	LPIPS	0,325	0,324	0,324	0,324	0,414	0,414	0,413	0,412	

		Cylindical + Equirectangular				Spherical				
		SH 0	SH 1	SH 2	SH 3	SH 0	SH 1	SH 2	SH 3	Threshold
Fields	SSIM	0,615	0,615	0,614	0,611	0,615	0,615	0,614	0,612	80
	PSNR	19,80	19,82	19,73	18,94	19,73	19,69	19,60	19,40	
	LPIPS	0,290	0,289	0,290	0,293	0,302	0,302	0,303	0,305	

		Cylindical + Equirectangular				Spherical				
		SH 0	SH 1	SH 2	SH 3	SH 0	SH 1	SH 2	SH 3	Threshold
Plane	SSIM	0,554	0,553	0,550	0,547	0,551	0,549	0,546	0,542	40
	PSNR	16,10	16,07	16,02	15,92	15,92	15,89	15,81	15,69	
	LPIPS	0,415	0,415	0,415	0,416	0,423	0,423	0,425	0,426	

		Cylindical + Equirectangular				Spherical				
		SH 0	SH 1	SH 2	SH 3	SH 0	SH 1	SH 2	SH 3	Threshold
Francis	SSIM	0,706	0,704	0,701	0,696	0,709	0,708	0,704	0,698	20
	PSNR	17,38	17,21	16,89	16,42	17,97	17,89	17,70	17,15	
	LPIPS	0,338	0,341	0,345	0,349	0,362	0,365	0,369	0,376	

Table 3: Quantitative evaluation of novel view synthesis replacing background Gaussians with environment maps. We compare our Cylindrical Slit-Scan method against a direct Spherical rendering approach, evaluating different Spherical Harmonics (SH) bands. As expected, the lack of motion parallax in static 2D maps inherently penalizes pixel-wise metrics (PSNR, SSIM) compared to full 3D representations, though perceptual stability in VR remains stable. Moreover, LPIPS results are less sensitive to these problems. Our Cylindrical method generally outperforms or matches the Spherical one while avoiding severe geometric distortions at the poles.