Smartphone Bi-Camera System for Backwards Photometric Correction of a 3D Gaussian Splatted Render

Rhea Joyce Zambra[a], Hongzhou Yang[a]

[a]Dept. of Geomatics Engineering, University of Calgary, Alberta, Canada – honyang@ucalgary.ca

KEY WORDS: Novel View Synthesis, Gaussian Splatting, Photometric Error, Multi-Camera

ABSTRACT:

Multi-camera systems are essential for multi-view stereo, enabling accurate depth estimation and 3D reconstruction through parallax. However, traditional bi-camera setups, typically forward-facing, are limited in capturing diverse environmental perspectives, particularly in autonomous driving, where vehicle motion is restricted. To overcome these limitations, we propose a novel no-parallax bi-camera system that combines a forward-facing and a backward-facing camera on a smartphone setup. This configuration simulates 360° spatial coverage and enhances 3D reconstruction by photometrically correcting forward-facing Gaussian-splatted renders using the backward-facing camera in a loosely-coupled manner, rather than relying on multi-view images in a conventional bundle adjustment. Our results show that incorporating a backward-facing image significantly improves the quality of backward renders, effectively reducing artifacts and mitigating overfitting seen in forward-facing-only images, resulting in a more accurate and refined 3D representation of the environment.

1. INTRODUCTION

Multi-camera systems play a critical role in multi-view stereo, enabling precise depth estimation and 3D scene reconstruction through parallax between overlapping fields of view. However, traditional bi-camera configurations face inherent limitations, such as requiring significant overlap between the cameras, which are typically oriented in the same forward direction. In the context of autonomous driving and outdoor mapping, this challenge is further exacerbated by the non-holonomic nature of the vehicle's motion, which is primarily constrained to a forward trajectory. As a result, the system's capacity to capture diverse perspectives of the environment from differing camera poses is significantly restricted. This limitation hinders the camera system's ability to create an accurate and comprehensive 3D representation of driving scenes.

Fundamentally, an effective camera network with convergent perspectives is essential for accurate and robust 3D reconstruction. Thus, modern autonomous systems often integrate multi-camera setups with varying orientations, sensor fusion with LiDAR or other depth sensors, and advanced perception algorithms in their systems. While these methods enhance scene understanding, they often require labour-intensive spatiotemporal calibration of the sensor suite, presenting challenges in terms of accessibility, scalability, and ease of integration, in addition to cost. Though monocular and stereo camera setups offer cost-effectiveness and rich visual information, the typical forward-facing setups often experience increased drift and reduced odometry accuracy compared to 360° LiDAR systems, which have superior spatial coverage (Agostinho *et al.*, 2022). Recent advances in novel view synthesis (NVS) techniques offer the potential to simulate a diverse camera network from image-only inputs. Notable approaches include Neural Radiance Fields (NeRF) (Mildenhall *et al.*, 2020), which employs implicit neural representations, and 3D Gaussian Splatting (3DGS) (Kerbl *et al.*, 2023) which utilizes explicit 3D Gaussian primitives. 3DGS has achieved state-of-the-art performance in view-dependent, photorealistic 3D rendering, offering both high-quality results and fast rendering speeds. However, 3DGS depends heavily on the quality of the input images due to its single-view training, and is consequently susceptible to overfitting in scenarios with limited training data, such as few-shot learning (Chen and Wang, 2024) or sparse-view configurations (Wu *et al.*, 2024). Thus, 3DGS often struggles to generalize to novel viewpoints that deviate significantly from the input trajectory, such as backward-facing views. Artifacts and reduced photometric accuracy are common in these scenarios, as illustrated in Figure 1.

To address these challenges while maintaining feasibility, we propose a smartphone bi-camera system with one forward-facing and one backward-facing camera. The backward-facing camera serves to photometrically correct the Gaussian-splatted render of the forward trajectory, mitigating artifacts and enhancing reconstruction accuracy. This setup provides an efficient, lightweight solution that simulates 360° spatial coverage, making it well-suited for autonomous perception.

Figure 1: Forward (top) and backward (bottom) Gaussian Splatting render of a KITTI image set, showing poor photorealism in the backwards render.

2. RELATED WORK

Different approaches have been proposed to address the limitations of Gaussian splatting in generalization from limited inputs, with key challenges including the randomly ordered single-view training and sensitivity to the structure-from-motion (SfM) initialization. Research in improving 3DGS NVS are discussed in the following subsections and are typically evaluated through photometric accuracy metrics such as peak signal-to-noise ratio (PSNR ↑), structural similarity index (SSIM ↑), and learned perceptual image patch similarity (LPIPS ↓).

2.1 Machine Learning and Related Approaches

Recent advancements in 3D Gaussian Splatting (3DGS) have sought to enhance generalizability through sophisticated machine learning techniques. For instance, Self-Ensembling Gaussian Splatting (Zhao *et al.*, 2024) introduces a self-ensembling model that aggregates information from samples with varying uncertainty-based perturbations, serving as an effective regularization mechanism. In a similar but stripped-down approach, CoR-GS (J. Zhang *et al.*, 2024) utilizes co-regularization by simultaneously training two 3DGS models and leveraging the differences from randomness in densification. PixelSplat (Charatan *et al.*, 2024) tackles the propagation of initialization errors by using a neural network to predict a dense 3D probability distribution, from which 3D Gaussians are sampled rather than directly placed. This probabilistic approach stabilizes training and improves the robustness of 3D scene reconstruction. Meanwhile, TranSplat (C. Zhang *et al.*, 2024) uses a transformer-based architecture for sparse-view scene reconstruction, combining monocular depth estimation with a Depth Refine U-Net and MLP to predict 3D Gaussian parameters (position, opacity, covariance, color). While these approaches effectively use internally derived information to mitigate distortions in viewpoints similar to the input images, challenges remain in handling extreme viewpoint deviations for 3DGS renders. Additionally, computational load, black-box architecture, and large dataset requirements continue to pose significant barriers to scalability and real-time applications.

2.2 Geometric Prior Approaches

Geometric priors have been increasingly utilized to enhance the robustness of 3DGS. Multi-view geometry techniques, in particular, have gained significant traction as a means of counteracting the overfitting associated with the single-view training of 3DGS models. Since vanilla 3DGS relies on image-only input, traditional photogrammetric methods and machine learning models are commonly employed to extract correspondences between multiple views, as outlined below.

Depth regularization is a popular paradigm in improving 3DGS NVS, with methods such as DNGaussian (J. Li *et al.*, 2024), CoherentGS (Paliwal *et al.*, 2025), and MVSGaussian (Liu *et al.*, 2024) using machine-learning aided monocular depth estimation to jointly optimize depth and the 3D Gaussians from the scene. On the other hand, Multi-View Regulated Gaussian Splatting (MVGS) (Du, Wang and Yu, 2024) enhances 3D Gaussian optimization by combining the gradients from camera intrinsics, camera extrinsics, and 3D Gaussian parameters across multiple views within a single loss function. This approach enables the simultaneous refinement of different images in a single iteration, improving consistency and rendering quality. Similarly, MVSplat (Chen *et al.*, 2025) aggregates multi-view information through transformer-based feature extraction and 2D U-Net depth recovery, which are then projected to characterize the 3D Gaussians. DynSUP (W. Li *et al.*, 2024) and Splatt3R (Smart *et al.*, 2024) produce high-fidelity 3D renders from unposed image pairs, with the former using traditional dense bundle adjustment (BA) at an object-level, and the latter using cross-attention transformers to derive correspondences. Although these approaches improve photometric accuracy (as measured by PSNR, SSIM, and LPIPS), they are mostly evaluated in datasets with full six-degree-of-freedom (6-DoF) movement or highly convergent rays, which do not fully reflect the motion constraints in autonomous driving scenarios (Table 1). Furthermore, the use of bi-camera systems for Gaussian Splatting in multi-view reconstruction remains underexplored, highlighting the need for a more practical, easily integrable, and computationally efficient solution for non-holonomic scenarios.

2.3 Proposed Bi-Camera System

To overcome these challenges, we propose a novel bi-camera system consisting of two smartphones, one facing forward and one facing backwards. This configuration eliminates overlap between the cameras' fields of view, forming a loosely-coupled bi-camera system where the backward-facing camera is not used for triangulation, but rather to photometrically correct the renders generated by the forward-facing camera. Our contribution lies in exploring the potential of no-parallax, bidirectional setups for enhancing Neural View Synthesis (NVS) using Gaussian Splatting in the context of autonomous driving. By utilizing this complementary data, our approach offers a more data-efficient and computationally feasible solution for improved NVS, which can be seamlessly integrated into existing systems.

Dataset Name	Camera Description
Mip-NeRF 360	Single camera, rotating around an object
Tanks and Temples	Single camera, 6DOF
Deepblending	Single camera, 6DOF
DTU	Single camera, 6DOF
NeRF Synthetic	Single camera, close-bounded with circular movements
NVS-RGBD	Single camera, close-bounded with circular movements

Table 1. Popular datasets used for photometric accuracy assessments for NVS.

3. PRELIMINARY

3.1 Structure from Motion through COLMAP

Gaussian splatting intakes a set of posed images $\mathcal{I} = \{\mathcal{I}_i \mid i = 1, \dots, N_I\}$ and a sparse set of 3D points $\mathcal{X} = \{X_k \in \mathbf{R}^3 \mid k = 1 \dots N_X\}$ which is obtained from a SfM adjustment through COLMAP (Schönberger and Frahm, 2016), which accepts an unordered set of images. First, feature extraction for each image is performed, resulting in correspondences $\mathcal{C}$ between image pairs I_a and I_a:

$$\mathcal{C} = \{\{I_a, I_b\} \mid I_a, I_b \in \mathcal{I}, a < b\} \quad (1)$$

Then after a two-view reconstruction, images are registered through a perspective-n-point (PnP) algorithm, obtaining the poses P_c and camera intrinsics, if uncalibrated.

$$\mathcal{P} = \{P_c \in SE(3) \mid c = 1 \dots N_P\} \quad (2)$$

Registered images must observe existing points and new scene points, which are added through triangulation. After a threshold, a bundle adjustment is performed to minimize the reprojection error, to prevent the reconstruction from drifting to a non-recoverable state:

$$E = \sum_j \rho_j \left(\left\| \pi(P_c, X_k) - x_j \right\|_2^2 \right) \quad (3)$$

where π is a perspective transformation, x_j is the 2D image point, and ρ_j is a loss function that reduces the influence of outliers.

3.2 3D Gaussian Splatting

In 3D Gaussian Splatting (3DGS), each point is represented by a 3D Gaussian primitive, meaning it has a spatial extent, orientation, and shape. This helps in representing continuous surfaces rather than just discrete points. To describe the size, shape, and orientation of these Gaussians, each one is characterized by: a) a covariance matrix Σ, which controls scale and spread in all directions; b) a mean μ, which controls the position; c) spherical harmonic coefficients C, which control the color; and d) opacity α. The 3D Gaussian function on point x in space can be represented as:

$$G(x) = e^{-\frac{1}{2} x^T \Sigma^{-1} x} \quad (4)$$

which results in a 3D volumetric and continuous representation of a scene. The covariance Σ is parametrized as:

$$\Sigma = RSS^T R^T \quad (5)$$

to enable optimization using gradient descent and anisotropy of the Gaussians, which allows for high-fidelity representation of complex areas (many anisotropic 3D Gaussians) and compact representation of sparse scenes (fewer, larger 3D Gaussians). The Gaussians are initially initialized as isotropic points derived from the Structure-from-Motion (SfM) sparse point cloud. Tile-based rasterization is then performed by dividing the screen into tiles, and in each tile the Gaussians are ordered by depth. During the forward pass, color and α are accumulated per pixel as Gaussians are blended front-to-back (α-blending). For backpropagation, the four parameters are optimized by comparing the render to the ground truth images, using a loss function that combines L1 (absolute difference of rendered and true pixel value) and D-SSIM (preserve higher-level structure of photo) for photorealism:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda \mathcal{L}_{D-SSIM} \quad (6)$$

where $\lambda = 0.2$ is a hyperparameter that controls the balance between the two loss terms. Additionally, Gaussian Splatting implements adaptive density control, where it clones 3D Gaussians in underrepresented parts of the scenes and splits them in over-reconstructed areas. It also culls Gaussians that are below an opacity threshold to prevent artifacts.

Currently, the original 3DGS source code employs a randomly ordered single-view training to optimize the parameters, which makes it susceptible to overfitting especially in non-holonomic trajectories where the images usually face only forward. Additionally, non-holonomic monocular trajectories can exhibit significant drift when inputted into SfM techniques due to low parallax between subsequent images (Zambra *et al.*, 2024) for high and constant frame-rate systems, such as those used in autonomous vehicles.

4. METHODOLOGY

4.1 Data Collection

Two Google Pixel 7 smartphones (GP7P1 – forward, GP7P4 – backwards) were used for this experiment, capturing images alongside UNIX timestamps via the SensorLogger app (Tszheichoi, 2024) at a frame rate of 30 FPS (Figure 2). The phones were mounted side by side on a cart with a short, uncalibrated baseline (Figure 3), and data collection took place indoors at the Schulich School of Engineering, University of Calgary using a forward-only trajectory with turns (Figure 4). Each smartphone underwent two data collection sessions at the 2nd (ENGG2) and 4th (ENGG4) floors of the building, resulting in a total of four image sets which were downsampled (Table 2) to evaluate the proposed bi-camera system. The images were cropped to omit static parts of the photo (ie. the cart). Lastly, to achieve coarse time synchronization, images taken before the ego-cart was in motion were discarded.

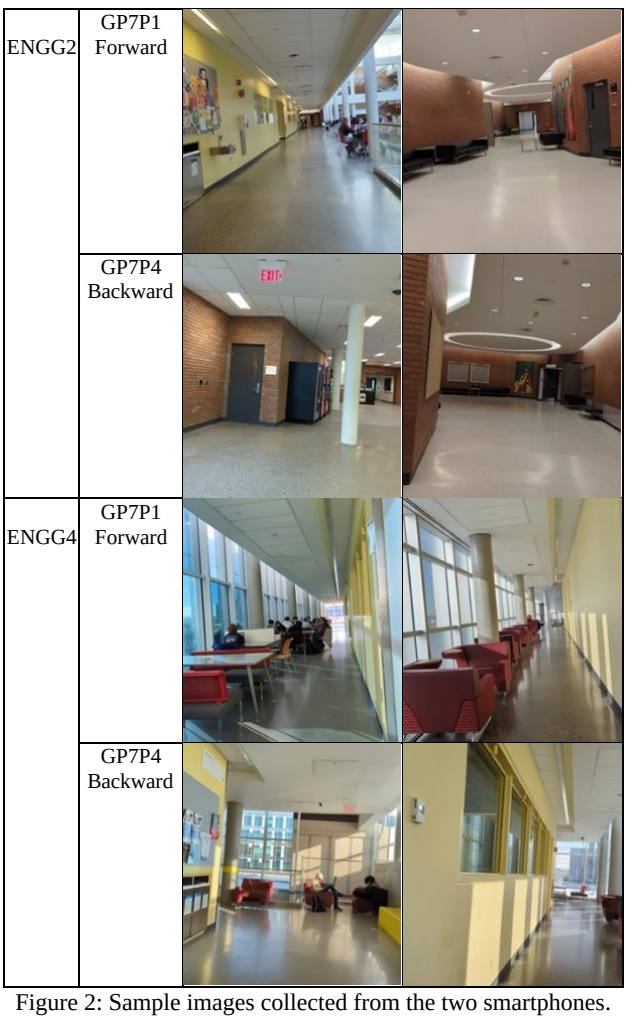

Figure 2: Sample images collected from the two smartphones.

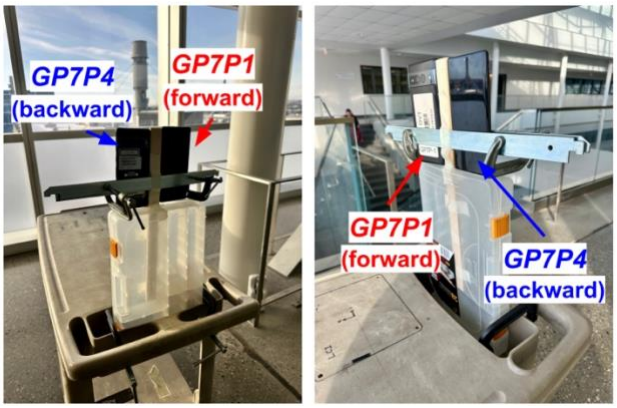

Figure 3: The two-smartphone platform used for data collection.

	ENGG2		ENGG4	
	Forward	**Backward**	**Forward**	**Backward**
Number Images	2358	262	653	14
Description	With loops, feature-rich		No loop corridor, with one turn	
Resolution	2125 × 2367 [pixels]			

Table 2: Metadata of each smartphone image set.

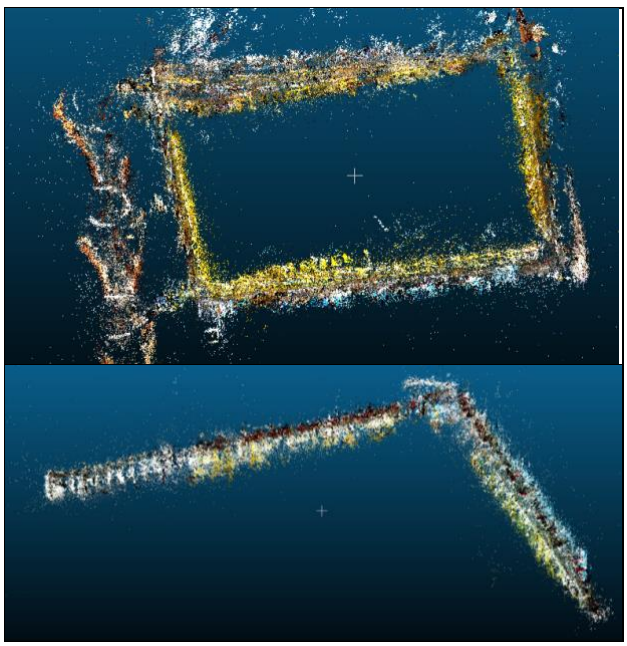
Figure 4: The indoor trajectories for ENGG2 (top) and ENGG4 (bottom).

4.2 Data Preprocessing

To mitigate reconstruction drift, forward-facing images were downsampled by a factor of four (resulting in 7.5 FPS) before being processed in COLMAP to extract image extrinsics and generate a sparse point cloud and prevent drift. The backwards facing images were downsampled at a higher rate to prevent artifacts from spawning in the forward-facing render. The original 3DGS implementation includes a COLMAP converter that outputs an *image.txt* file, which contains the image index, image filename, camera pose (represented as a quaternion and a 3D translation vector), and 2D-3D correspondences. A corresponding */image* directory stores all images used for 3DGS training. In this experiment, the methodology is summarized as:

1. **Downsample backward-facing images:** A downsampling factor of 30 was heuristically found to be optimal for ENGG2 (a feature-rich environment with loops), while a factor of 90 worked best for ENGG4 (a straight corridor without loops).
2. **Establish temporal correspondence:** After time synchronization, a mapping table was created to link forward-facing image timestamps (or indices) to their corresponding backward-facing images.
3. **Integrate backward-facing images:** Using the COLMAP reconstruction, forward-facing images with corresponding backward-facing images were identified. The backward-facing images were then appended to image.txt with their poses locally rotated by 180°.
4. **Update the image dataset:** The matched backward-facing images were added to the */images* directory for training.
5. **Train and compare reconstructions:** Two separate 3DGS training runs were conducted—one using only forward-facing images and another using both forward- and backward-facing images—to evaluate the impact of additional viewpoints on reconstruction quality.

The baseline offset was not accounted for in data augmentation due to: (a) the inherent scale ambiguity in monocular COLMAP reconstructions, and (b) the flexibility of the four trainable parameters in 3DGS, which require only approximate initial values rather than exact ones.

5. RESULTS

The results show similar performance for the forward-facing renders (F1, F2, F3), with a minor loss of detail, particularly on glossy surfaces. However, a notable improvement is observed in the backward-facing render (B1, B2, B3) when both image sets are used across both data collection areas (Figures 5-6). The PSNR and SSIM metrics are summarized in Table 3.
Qualitatively, the inclusion of backward-facing images acts as a 'smoothing' mechanism, preventing the 3D Gaussians from becoming excessively sharp and anisotropic, which could otherwise lead to overfitting to the forward-facing images. Furthermore, bounded structures, such as pillars, are more accurately represented due to the constraints provided by both the forward and backward views. An especially noteworthy case occurs when turning a corner in ENGG4 (Figure 6, row 5), where the forward-only render is completely disintegrated, likely due to the drift in the camera extrinsic estimation while it is rotating. In contrast, using both image sets maintains photorealism, with the sofa remaining clearly distinguishable.

	ENGG2				ENGG4			
	F		FB		F		FB	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
F1	28.161	**0.659**	**28.196**	0.656	28.161	**0.659**	**28.196**	0.656
F2	29.785	**0.736**	**31.387**	0.712	28.872	**0.707**	**29.292**	0.703
F3	**28.539**	0.690	28.134	**0.716**	28.337	**0.660**	**28.375**	0.641
B1	27.882	0.705	**28.321**	**0.748**	27.935	0.687	**28.117**	**0.724**
B2	28.298	0.681	**28.536**	**0.689**	27.935	0.687	**28.254**	**0.703**
B3	28.072	0.644	**29.015**	**0.684**	27.937	0.620	**28.593**	**0.684**

Table 3: A comparison of the PSNR (↑) and SSIM(↑) for forward-only (F) vs. forward-and-backward renders (FB).

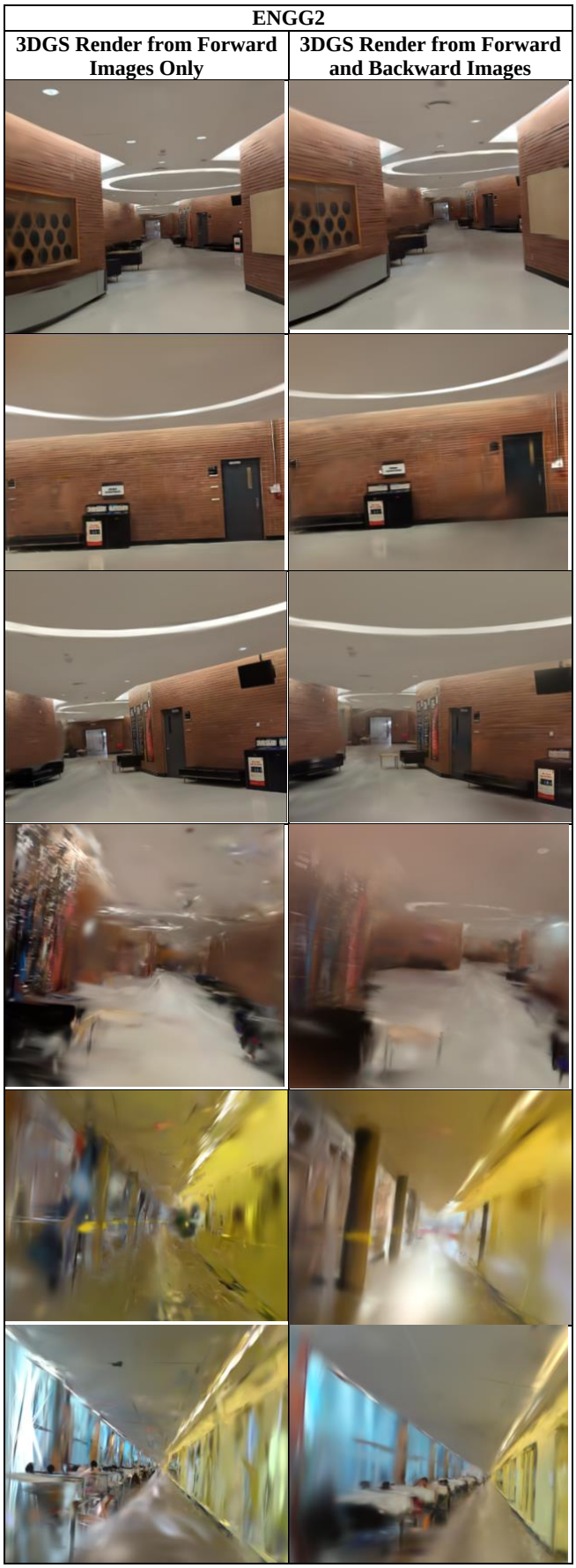

Figure 5: Comparison of backward renders from the Forward-Only and Forward-and-Backward 3DGS renders for ENGG2.

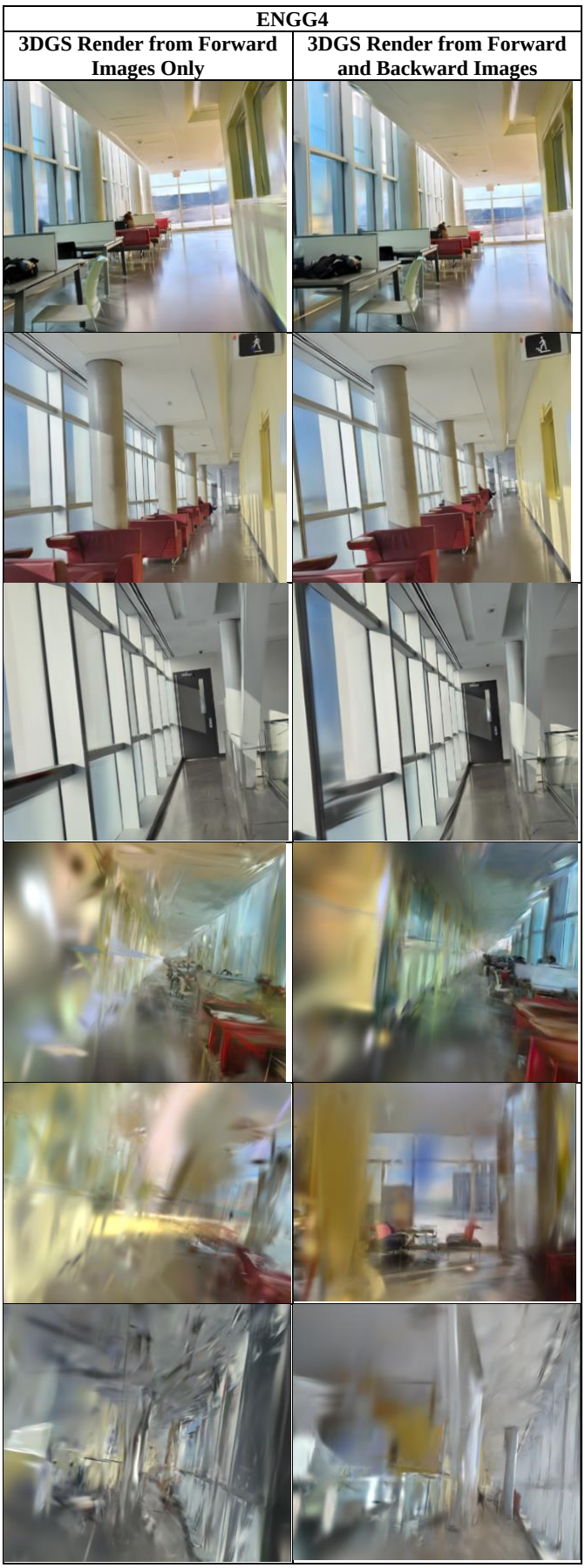

Figure 6: Comparison of the backward renders from Forward-Only and Forward-and-Backward 3DGS renders for ENGG4.

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-1677-2025

Additionally, a comparison of the 3D point clouds generated from 3DGS densification reveals that the forward-only model is more prone to artifacts (Figures 7-8). This leads to an excess of Gaussians acting as "embellishments" to overfit the forward-facing view (Table 4).

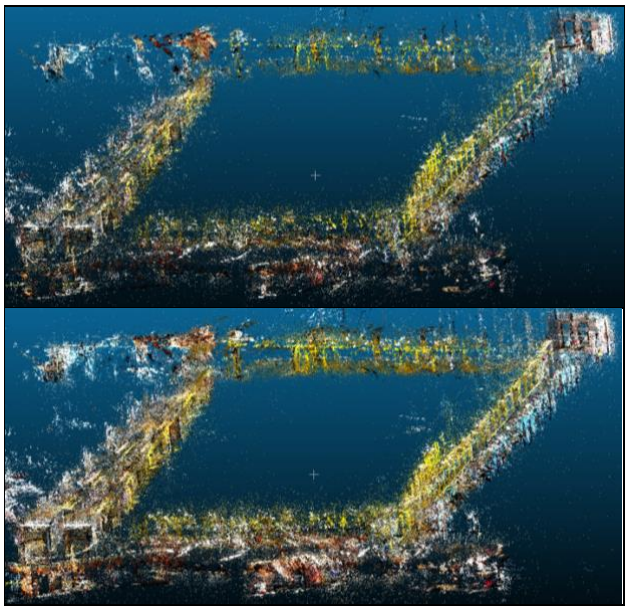

Figure 7: Comparison of 3DGS point clouds: forward-and-backward results (top), and overlay of forward-only and forward-and-backward (bottom) for ENGG2.

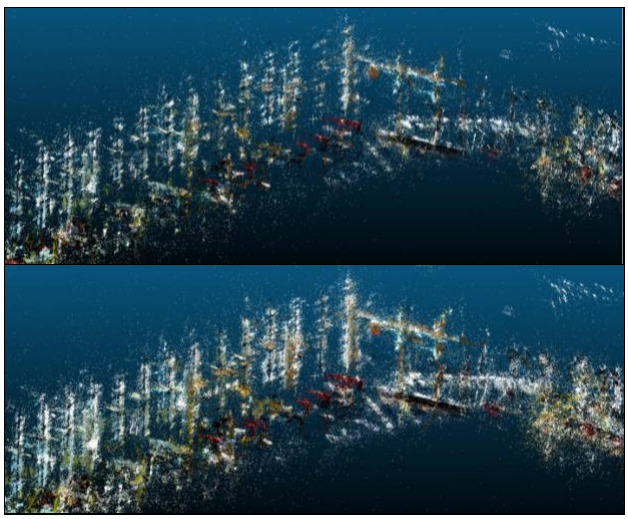

Figure 8: Comparison of 3DGS point clouds: forward-and-backward results (top), and overlay of forward-only and forward-and-backward (bottom) for ENGG4.

	ENGG2		ENGG4	
	Forward Only	**Forward-Backward**	**Forward Only**	**Forward-Backward**
Number of Points	670,962	444,957	853,692	833,626
Number of Images	2,358	2,620	653	667

Table 4: The number of points contained in each resulting 3DGS point cloud.

In the closed loop environment of ENGG2, the forward only point cloud contained 50% more points than the forward-and-backward render, showing that more points may not always be desired and can indicate overfitting. This demonstrates that quality in 3D reconstructions is not solely determined by the quantity of points, but by how effectively they represent the underlying scene. This finding underscores the importance of accounting for additional views to improve model generalization and reduce overfitting.

6. CONCLUSIONS AND OUTLOOK

In this paper, we have explored the limitations of traditional bi-camera systems and the challenges posed by single-view training in 3D Gaussian Splatting (3DGS) for Neural View Synthesis (NVS). We introduced a novel bi-camera configuration utilizing forward- and backward-facing smartphone cameras to enhance 3DGS reconstruction accuracy. Our approach leverages the complementary nature of the backward-facing camera, not for triangulation but for photometric correction, effectively mitigating artifacts and improving rendering consistency. By addressing the limitations of traditional forward-facing camera setups, our method provides an efficient and lightweight alternative to existing multi-view stereo and sensor fusion techniques, making it particularly well-suited for autonomous perception in constrained vehicular motion scenarios.

Our results demonstrate that integrating a backward-facing camera into a smartphone-based system can significantly enhance NVS performance, especially in scenarios with limited training data or sparse-view configurations. The proposed method offers a cost-effective and scalable solution that minimizes the need for complex calibration while maintaining high reconstruction fidelity. Furthermore, by reducing reliance on computationally intensive machine learning models, our approach enhances real-time applicability and accessibility for autonomous driving systems and outdoor mapping applications.

Future work in this area includes assessing the 3D reconstruction in addition to photometric accuracy, investigating the use of alternative viewpoints, such as side-facing or top-down configurations, and evaluating the system in outdoor driving scenarios.

Acknowledgements

We gratefully acknowledge the support of Alberta Innovates in this project.

References

Agostinho, L.R. *et al.*, 2022. 'A Practical Survey on Visual Odometry for Autonomous Driving in Challenging Scenarios and Conditions', *IEEE Access*, 10, pp. 72182–72205.

Charatan, D., Li, S.L., Tagliasacchi, A. and Sitzmann, V., 2024. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19457-19467.

Chen, G. and Wang, W., 2024 'A Survey on 3D Gaussian Splatting'. *arXiv preprint arXiv:2401.03890*.

Chen, Y. *et al.*, 2025. 'MVSplat: Efficient 3D Gaussian Splatting from Sparse Multi-view Images', in A. Leonardis et al. (eds) *Computer Vision – ECCV 2024*. Cham: Springer Nature Switzerland, pp. 370–386.

Du, X., Wang, Y. and Yu, X., 2024. Mvgs: Multi-view-regulated gaussian splatting for novel view synthesis. *arXiv preprint arXiv:2410.02103*.

Kerbl, B. *et al.*, 2023. ‘3D Gaussian Splatting for Real-Time Radiance Field Rendering’, *ACM Transactions on Graphics*, 42(4), p. 1.

Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J. and Gu, L., 2024. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20775-20785).

Li, W., Chen, W., Qian, S., Chen, J., Cremers, D. and Li, H., 2024. DynSUP: Dynamic Gaussian Splatting from An Unposed Image Pair. *arXiv preprint arXiv:2412.00851*.

Liu, T., Wang, G., Hu, S., Shen, L., Ye, X., Zang, Y., Cao, Z., Li, W. and Liu, Z., 2024, September. Mvsgaussian: Fast generalizable gaussian splatting reconstruction from multi-view stereo. In *European Conference on Computer Vision* (pp. 37-53). Cham: Springer Nature Switzerland.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), pp.99-106.

Paliwal, A., Ye, W., Xiong, J., Kotovenko, D., Ranjan, R., Chandra, V. and Kalantari, N.K., 2024, September. Coherentgs: Sparse novel view synthesis with coherent 3d gaussians. In *European Conference on Computer Vision* (pp. 19-37). Cham: Springer Nature Switzerland.

Schonberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. In Proceedings of the *IEEE conference on computer vision and pattern recognition* (pp. 4104-4113).

Smart, B., Zheng, C., Laina, I. and Prisacariu, V.A., 2024. Splatt3r: Zero-shot gaussian splatting from uncalibrated image pairs. *arXiv preprint arXiv:2408.13912*.

Tszheichoi, 2024. SENSOR LOGGER, Version 1.31.3. Kelvin Choi. https://www.tszheichoi.com/sensorlogger (1 April 2024).

Wu, T., Yuan, Y.J., Zhang, L.X., Yang, J., Cao, Y.P., Yan, L.Q. and Gao, L., 2024. Recent advances in 3d gaussian splatting. *Computational Visual Media*, 10(4), pp.613-642.

Zambra, R.J., Uprety, S., Lee, R. and Yang, H., 2024, September. Smartphone HD Map Updates Using Monocular-Inertial ORB-SLAM3 and Gaussian Splatting. In Proceedings of the *37th International Technical Meeting of the Satellite Division of The Institute of Navigation* (ION GNSS+ 2024) (pp. 1885-1900).

Zhang, C., Zou, Y., Li, Z., Yi, M. and Wang, H., 2024. Transplat: Generalizable 3d gaussian splatting from sparse multi-view images with transformers. *arXiv preprint arXiv:2408.13770*.

Zhang, J., Li, J., Yu, X., Huang, L., Gu, L., Zheng, J. and Bai, X., 2024, September. CoR-GS: sparse-view 3D Gaussian splatting via co-regularization. In *European Conference on Computer Vision* (pp. 335-352). Cham: Springer Nature Switzerland.

Zhao, C., Wang, X., Zhang, T., Javed, S. and Salzmann, M., 2024. Self-Ensembling Gaussian Splatting for Few-Shot Novel View Synthesis. *arXiv preprint arXiv:2411.00144*.