IncEventGS: Pose-Free Gaussian Splatting from a Single Event Camera

Supplementary Material

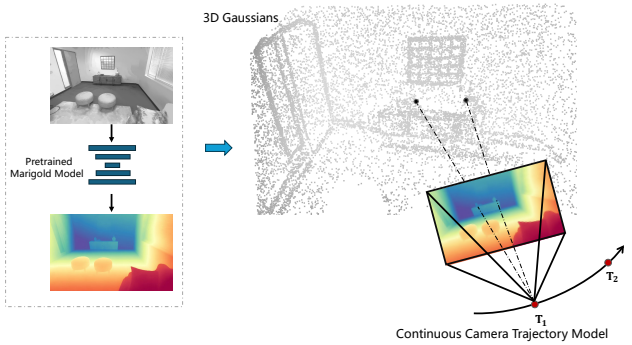

Figure 1. The re-initialization process of *IncEventGS*.

1. More Details about Re-initialization

The re-initialization process is illustrated in Fig. 1. After the first-time initialization, we can render a brightness image from 3D-GS at pose $\mathbf{T}_1$, where $\mathbf{T}_1$ represents the camera pose at the end of the first event chunk. To improve the 3D structure of 3D-GS, we use a monocular depth estimation network [2] to predict a dense depth map from the rendered brightness image. This depth map is then used to re-initialize the centers of the 3D Gaussians by unprojecting the pixel depths at camera pose $\mathbf{T}_1$, as illustrated in Fig. 1. After re-centering the 3D Gaussians, we perform the initialization process again to achieve both accurate 3D structure and exceptional brightness image rendering performance.

2. Comparison with Gaussian-based Event Methods

To further evaluate our method, we conducted additional comparisons against state-of-the-art Gaussian-based event approaches. Since Event3DGS [7] had not been open-sourced, we chose to compare against E2GS[1] and EvGGS[6]. In particular, we removed the supervision of blurred image in E2GS and exploited the pretrained weight of EvGGS for comparisons. As shown in Table 2, our method still outperforms those two baselines event though they used ground truth poses. Since EvGGS is a generalizable method based on a feed-forward network, it has limited generalization capability on unseen dataset.

3. Experiments in Fast-Motion Scenarios

Fast camera movement can induce motion blur, making it challenging to reconstruct the scene and estimate camera poses using RGB-based algorithms. We compare our event-based method with two state-of-the-art pose-free Gaussian SLAM implementations: MonoGS [4] (RGB modality) and SplaTAM [3] (RGBD modality). By leveraging the high temporal resolution of event cameras, our method experiences minimal performance degradation, even under fast motion. Additionally, it is more effective at preserving high-frequency information in the scene. As shown in Fig. 2, our approach delivers superior novel view synthesis results, particularly during rapid camera movement.

Method	Synthetic (768 × 480)		Real-world (1280 × 720)	
	Training	Storage	Training	Storage
ENeRF	12 hour	253M	12 hour	253M
EventNeRF	21 hour	14M	24 hour	14M
Robust e-NeRF	11 hour	745M	13 hour	745M
Ours	0.5 hour	65M	2 hour	55M

Table 1. Average model efficiency comparison.

4. Experiments on Color Event Datasets

Our method can also be applied to color event datasets by integrating the Bayer filter [5], as shown below:

$$\mathcal{L}_{event} = \left\| \mathbf{F} \odot \mathbf{E}_i(\mathbf{x}) - \mathbf{F} \odot \hat{\mathbf{E}}_i(\mathbf{x}) \right\|_2 \quad (1)$$

$$\mathcal{L}_{ssim} = SSIM(\mathbf{F} \odot \mathbf{E}_i(\mathbf{x}), \mathbf{F} \odot \hat{\mathbf{E}}_i(\mathbf{x})) \quad (2)$$

Furthermore, our method can be extended to incorporate training with ground-truth poses.

We conducted experiments on the EventNeRF dataset [5], which focuses on object reconstruction. Due to the dataset's limited features, pose estimation is challenging; neither COLMAP nor DEVO can estimate camera poses on this dataset. As shown in Fig. 3, our method can still successfully optimize both the 3D scene and camera poses even without ground-truth poses, though it produces minor artifacts. When trained with ground-truth poses, our method achieves improved novel view synthesis, with fewer artifacts and sharper textures.

5. Time Evaluations

As shown in Table 1, our method has a significant advantage in training time compared to NeRF-based methods. Additionally, our method achieves an NVS rendering speed of approximately 500 FPS, whereas NeRF-based methods reach only about 0.5 FPS.

We mainly focus on demonstrating the effectiveness (*i.e.* in terms of novel view synthesis and pose estimation) by exploiting 3D-GS representation for event camera, and have not tried to improve the efficiency of the proposed method. In

Figure 2. Qualitative evaluation of novel view image synthesis on the Replica dataset. The experimental results demonstrate that our method renders higher-quality images when the camera is moving fast.

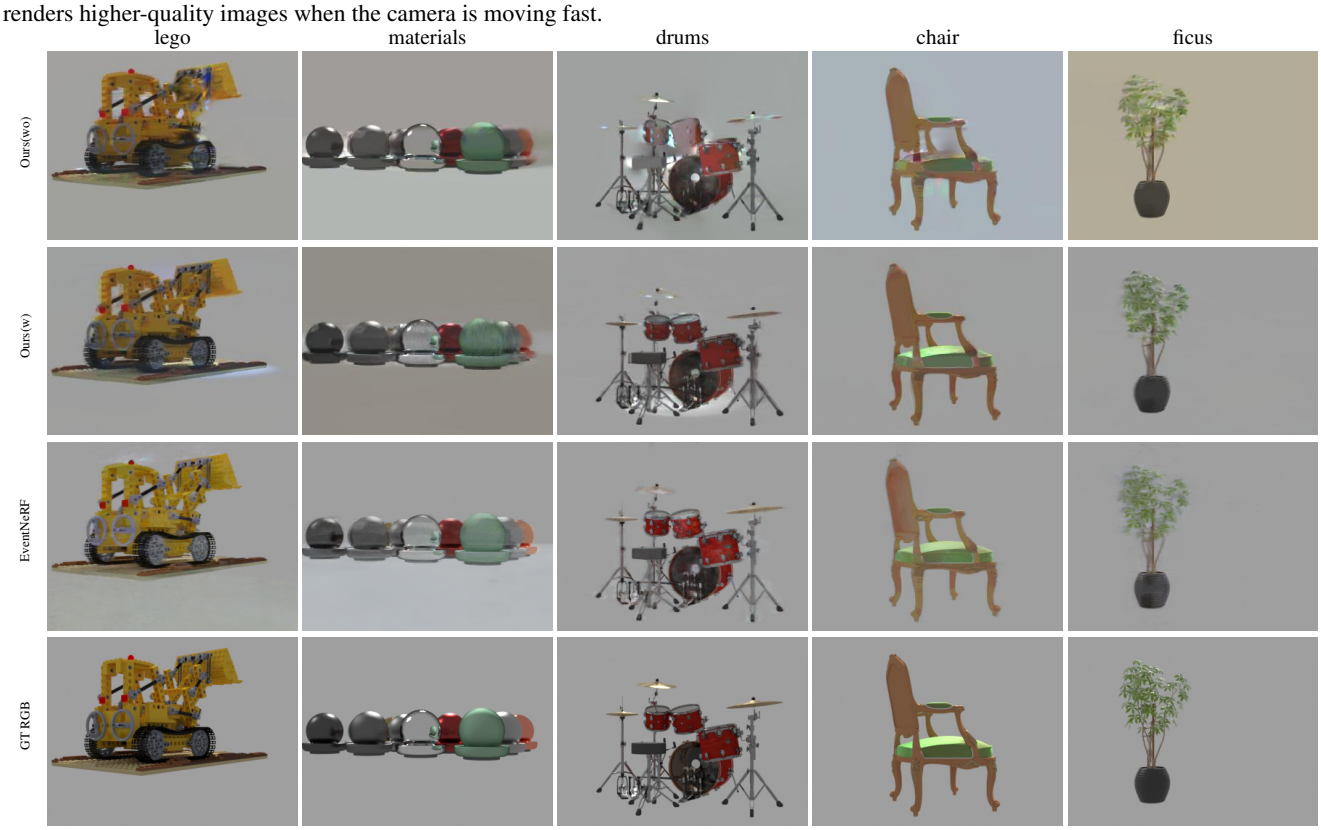

Figure 3. Qualitative evaluation of novel view image synthesis on color event dataset. Ours (wo) denotes our method trained without ground-truth camera poses, while Ours (w) denotes the method trained with ground-truth camera poses.

	room0			*room2*			*office0*			*office2*			*office3*		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
E2GS*	21.75	0.77	0.25	23.11	0.82	0.20	20.09	0.75	0.18	18.62	0.78	0.20	20.13	0.84	0.16
EvGGS	15.16	0.37	0.62	15.85	0.34	0.61	18.51	0.37	0.59	10.95	0.27	0.69	13.13	0.29	0.66
Ours	**24.31**	**0.85**	**0.17**	**23.75**	**0.79**	**0.23**	**25.64**	**0.54**	**0.30**	**21.74**	**0.82**	**0.23**	**21.18**	**0.88**	**0.13**

Table 2. NVS performance comparison on Replica dataset. * denotes we removed the supervision of blurred images from the original E2GS.The result demonstrates that our method outperforms those two baseline methods.

particular, for the ease of the development, we still adopt the Adam optimizer with a small learning rate (*i.e*. 1e-4) from PyTorch for both motion and 3D-GS estimation. It requires around 0.3s and 1.7s per event chunk to converge for both tracking and mapping respectively. We would further improve the efficiency by using a second-order optimization method (*e.g*. levenberg-marquardt algorithm), which has been proved to converge much faster to the optimal solution compared to an first-order optimizer (*e.g*. Adam).

References

[1] Hiroyuki Deguchi, Mana Masuda, Takuya Nakabayashi, and Hideo Saito. E2gs: Event enhanced gaussian splatting. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 1676–1682. IEEE, 2024. 1

[2] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1

[3] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat, track & map 3d gaussians for dense rgb-d slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 1

[4] Hidenobu Matsuki, Riku Murai, Paul HJ Kelly, and Andrew J Davison. Gaussian splatting slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18039–18048, 2024. 1

[5] Viktor Rudnev, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. Eventnerf: Neural radiance fields from a single colour event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4992–5002, 2023. 1

[6] Jiaxu Wang, Junhao He, Ziyi Zhang, Mingyuan Sun, Jingkai Sun, and Renjing Xu. Evggs: A collaborative learning framework for event-based generalizable gaussian splatting. *arXiv preprint arXiv:2405.14959*, 2024. 1

[7] Tianyi Xiong, Jiayi Wu, Botao He, Cornelia Fermuller, Yiannis Aloimonos, Heng Huang, and Christopher A Metzler. Event3dgs: Event-based 3d gaussian splatting for fast egomotion. *arXiv preprint arXiv:2406.02972*, 2024. 1