

# E-3DGS: Gaussian Splatting with Exposure and Motion Events

Xiaoting Yin<sup>1,\*</sup>, Hao Shi<sup>1,4,\*</sup>, Yuhan Bao<sup>1,\*</sup>, Zhenshan Bing<sup>4</sup>, Yiyi Liao<sup>3</sup>, Kailun Yang<sup>2,†</sup>, and Kaiwei Wang<sup>1,†</sup>

**Abstract**—Estimating Neural Radiance Fields (NeRFs) from images captured under optimal conditions has been extensively explored in the vision community. However, robotic applications often face challenges such as motion blur, insufficient illumination, and high computational overhead, which adversely affect downstream tasks like navigation, inspection, and scene visualization. To address these challenges, we propose E-3DGS, a novel event-based approach that partitions events into motion (from camera or object movement) and exposure (from camera exposure), using the former to handle fast-motion scenes and using the latter to reconstruct grayscale images for high-quality training and optimization of event-based 3D Gaussian Splattting (3DGS). We introduce a novel integration of 3DGS with exposure events for high-quality reconstruction of explicit scene representations. Our versatile framework can operate on motion events alone for 3D reconstruction, enhance quality using exposure events, or adopt a hybrid mode that balances quality and effectiveness by optimizing with initial exposure events followed by high-speed motion events. We also introduce EME-3D, a real-world 3D dataset with exposure events, motion events, camera calibration parameters, and sparse point clouds. Our method is faster and delivers better reconstruction quality than event-based NeRF while being more cost-effective than NeRF methods that combine event and RGB data by using a single event sensor. By combining motion and exposure events, E-3DGS sets a new benchmark for event-based 3D reconstruction with robust performance in challenging conditions and lower hardware demands. The source code and dataset will be available at <https://github.com/MasterHow/E-3DGS>.

## I. INTRODUCTION

Reconstructing 3D scenes and objects from images has been a fundamental research focus in various vision applications, including robotics navigation [1], [2], [3], virtual reality [4], [5], and scene understanding [6], [7], [8]. In robotic perception, accurate 3D reconstructions are vital for tasks like localization and path planning [1], [2], [3], while in virtual reality (VR) and augmented reality (AR), they underpin ego localization, spatial reasoning, and immersive

This was supported in part by Zhejiang Provincial Natural Science Foundation of China (Grant No. LZ24F050003), the National Natural Science Foundation of China (Grant No. 12174341 and No. 62473139), the National Key RD Program (Grant 2022YFB4701400), in part by Shanghai SUPREMIND Technology Co. Ltd, and in part by Hangzhou SurImage Technology Co. Ltd.

<sup>1</sup>State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, China (email: wangkaiwei@zju.edu.cn).

<sup>2</sup>School of Robotics and National Engineering Research Center of Robot Visual Perception and Control Technology, Hunan University, China (email: kailun.yang@hnu.edu.cn).

<sup>3</sup>College of Information Science and Electronic Engineering, Zhejiang University, China.

<sup>4</sup>The Chair of Robotics, AI, and Real-Time Systems, Technical University of Munich, Germany.

\*These authors contributed equally.

†Corresponding authors: Kaiwei Wang and Kailun Yang.

visualizations [4], [5]. Neural Radiance Field (NeRF) [9] has become a powerful tool for 3D reconstruction, but its limitations in training and rendering efficiency hinder real-time applications. 3D Gaussian Splattting (3DGS) [10] advances NeRF by using explicit 3D Gaussians and tile-based rasterization, enhancing both efficiency and synthesis quality.

Despite the advancements in NeRF [9] and 3DGS [10], it is notable that both methods heavily rely on high-quality training images to achieve accurate 3D reconstructions with RGB cameras [11]. Unfortunately, motion blur is a common challenge in real-world robotics systems, particularly in fast-moving scenarios or low-light conditions where longer exposure times are necessary. This blur often leads to mismatched feature points in COLMAP [12], resulting in inaccuracies in pose calibration and point cloud initialization, and sometimes even causing failures in recovering camera poses, thereby hampering the training of 3DGS.

Event cameras [13], which capture log-intensity changes asynchronously with microsecond-level resolution, provide a promising solution to these challenges. Several approaches [11], [14], [15], [16] have combined event cameras with NeRF to enhance 3D reconstruction in difficult environments. However, achieving real-time high-fidelity rendering remains a significant challenge for these implicit methods. Recent efforts [17], [18], [19] have explored integrating event cameras with 3D Gaussian Splattting (3DGS). However, these approaches overlook the source of event generation, often relying on additional RGB sensors for fusion or learning-based methods [20] to convert events into pseudo-grayscale images to achieve high-quality reconstructions, making it challenging to obtain such results using a single event sensor.

To achieve high-quality 3D reconstruction using only a single event sensor, we combine motion and exposure events to balance quality and efficiency in high-speed scenarios. We propose three operating modes for E-3DGS to address varying scene reconstruction needs:

- **High-Quality Reconstruction Mode:** In this mode, a programmable variable aperture gradually increases sensor brightness. The time each pixel reaches target brightness is recorded, generating grayscale images to guide 3DGS for precise low-light and high-dynamic-range reconstruction.
- **Fast Reconstruction Mode:** Utilizing a High-Definition (HD) resolution event camera to capture motion events, which are used to supervise 3D Gaussian Splattting (3DGS), enabling rapid 3D reconstruction in high-speed and low-light scenes.

TABLE I  
COMPARISON OF EVENT-BASED 3D RECONSTRUCTION DATASETS

| Feature          | EDS [21]   | EventNeRF [14] | EME-3D (Ours) |
|------------------|------------|----------------|---------------|
| Number of scenes | 16         | 7              | 9             |
| Image resolution | 346×260    | 346×260        | 1280×720      |
| Modality         | RGB, Event | RGB, Event     | Event         |
| Motion events    | ✓          | ✓              | ✓             |
| Exposure events  | ✗          | ✗              | ✓             |
| Number of images | 1,030      | 1,000          | 200           |
| Sharp images     | ✗          | ✓              | ✓             |
| Real data        | ✓          | ✗              | ✓             |

- **Balanced Hybrid Mode:** This mode starts with a few exposure events captured by the variable aperture, followed by motion events captured automatically. During 3DGS optimization, both exposure images and motion events jointly supervise the process, balancing reconstruction speed and quality in challenging scenarios.

Moreover, to validate the effectiveness of the proposed E-3DGS, we built a hardware acquisition system using a programmable variable aperture and a high-resolution event camera to create **EME-3D**, the first real-world 3D reconstruction dataset which distinguishes between **Exposure** and **Motion Events**. As shown in Tab. I, EME-3D contains nine sequences of 1280×720 high-resolution event streams, camera calibration parameters, sparse point clouds, and exposure events for reconstructing high-quality grayscale images. Experimental results demonstrate that, compared to event-based NeRF and event-to-grayscale learning-based 3DGS methods, E-3DGS achieves faster reconstruction, higher quality, and greater flexibility in handling diverse scene requirements. In Fast Reconstruction Mode, E-3DGS achieves a PSNR gain of 5.68dB over EventNeRF, along with a significantly higher rendering speed (79.37 FPS vs. 0.03 FPS). In the High-Quality Reconstruction Mode, E-3DGS delivers a PSNR increase of 10.89 dB compared to the event-to-grayscale learning-based 3DGS baseline.

The main contributions can be summarized as follows:

- We are the first to incorporate exposure event information into event-based 3D Gaussian Splatting (3DGS), converting sparse events during exposure into dense intensity frames for high-quality event-based 3D reconstruction.
- We establish and release EME-3D, the first real-world 3D dataset that contains both exposure and motion events, camera parameters, and sparse 3D point clouds.
- We propose a hybrid approach that captures exposure events at low speeds and motion events at high speeds, striking a good balance between reconstruction quality and efficiency.

## II. RELATED WORK

### A. Neural 3D Reconstruction

Traditional 3D reconstruction methods, such as point clouds [22], meshes [23], and voxel grids [24], [25], rely on explicit representations but are limited by their fixed topological structures. NeRFs [9] have gained traction for

synthesizing high-quality novel views using MLP-based neural networks and differentiable volume rendering, but they suffer from low rendering efficiency and long training times. More recently, 3D Gaussian Splatting [10] has demonstrated a significant advancement, offering faster convergence and superior rendering quality by employing Gaussian splats. However, all of these methods require clear RGB inputs and struggle with the fast motion or low-light conditions commonly encountered in real-world applications such as robotics and autonomous systems.

### B. Event-based 3D Reconstruction

Event cameras [13], which detect asynchronous brightness changes at the pixel level, offer substantial benefits in high dynamic range and temporally precise environments. Recent works have leveraged the complementary strengths of event streams and traditional RGB frames to enhance 3D scene reconstruction [11], [16], [19]. Incorporating event data into NeRF-based models [9], such as Ev-NeRF [15] and Event-NeRF [14], has demonstrated the potential for multi-view consistency in 3D reconstruction. However, NeRFs' computational demands hinder real-time rendering. To address this, event-based 3D Gaussian Splatting (3DGS) [17], [26] offers a more efficient alternative, accelerating reconstruction while preserving high-quality geometry and appearance. Nevertheless, due to limited texture in event streams, previous event-based 3DGS reconstructions often lack fine-grained detail [17], [26]. Therefore, some studies have attempted to overcome this by using learning-based event-to-image methods [18], [20], or event-image deblur methods [27], [28] to enhance the reconstruction accuracy. However, learning-based approaches require additional computational resources and yield limited quality improvements, while event-image modal fusion increases hardware complexity, size, and cost. In earlier work, EvTemMap [29] distinguished an event-based photography method by converting the precise timing of exposure events into dense grayscale frames, thereby eliminating the need for multiple sensors. Instead of introducing learning-based event-to-image or event-image deblur modules, we propose to explore exposure and motion events in 3D explicit reconstruction, which enhances both the quality and speed of reconstruction from pure event data, particularly beneficial in low-light or high-dynamic-range scenarios.

## III. METHODOLOGY

In this section, we present E-3DGS, an event-based method that partitions events into motion and exposure for high-quality training and optimization of event-based 3D Gaussian Splatting (3DGS), as illustrated in Fig. 1. Our approach utilizes the high temporal resolution of motion events and the rich texture information from exposure events, effectively integrating both capabilities within a single event camera. In the preliminary section (Sec. III-A), we introduce the 3DGS framework and the event camera model. Next, we explain how to map temporal information from exposure events into high-quality grayscale images (Sec. III-B) and

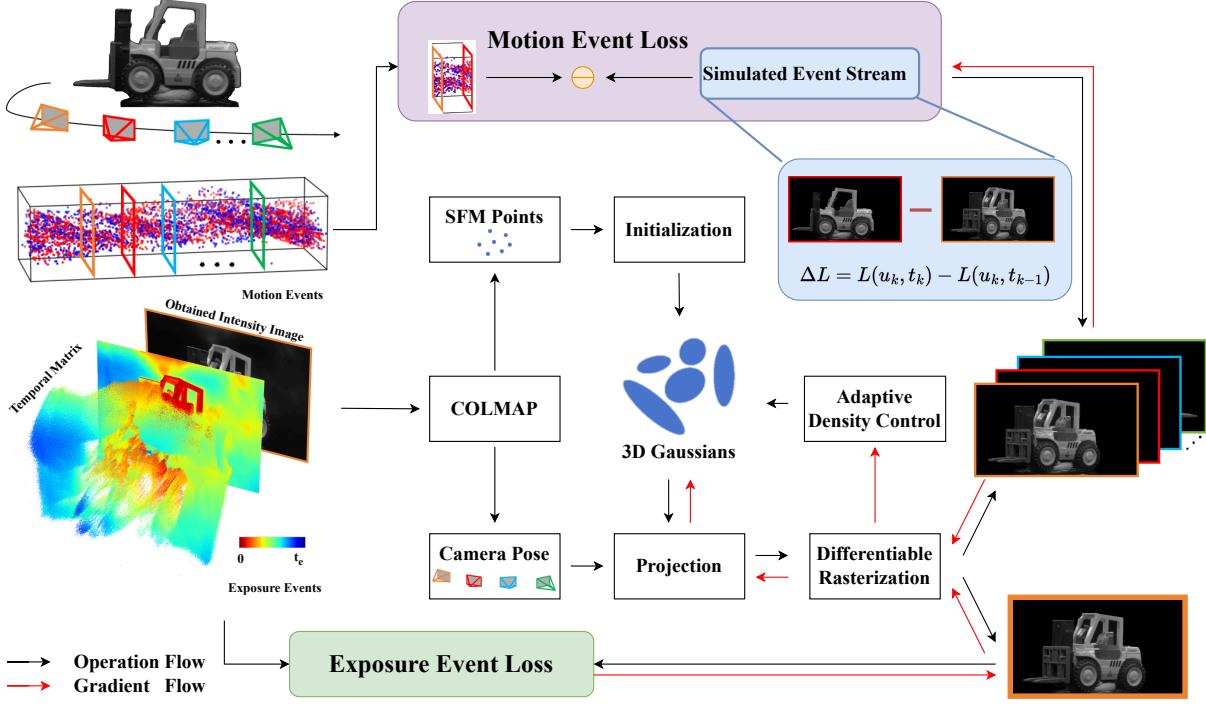


Fig. 1. Overview of the proposed E-3DGS framework. This framework integrates motion and exposure events for training 3DGS to effectively handle diverse real-world conditions. We utilize Temporal-to-Intensity Mapping to convert exposure events into intensity images, which yield camera trajectories and a sparse point cloud for 3DGS training. The optimization of 3DGS parameters is supervised through motion event loss and exposure event loss.

detailed the overall loss function in Sec. III-C. Finally, we describe the process for collecting real datasets in Sec. III-D.

#### A. Preliminary: 3D Gaussian Splatting & Event Model

In this section, we introduce the foundational elements of our approach, which combines 3D Gaussian Splatting (3DGS) and the event camera model.

**3D Gaussian Splatting (3DGS)** represents a 3D scene using anisotropic 3D Gaussians, each defined by a mean  $\mu \in \mathbb{R}^3$ , a covariance matrix  $\Sigma \in \mathbb{R}^{3 \times 3}$ , and an opacity  $\alpha$ . The covariance matrix  $\Sigma$  is factorized as:

$$\Sigma = RSS^\top R^\top, \quad (1)$$

where  $R$  is a rotation matrix and  $S$  is a scaling matrix, ensuring the matrix remains positive semi-definite for optimization. For rendering, the 3D Gaussians are projected onto the 2D image plane, with the covariance matrix transformed into camera coordinates as:

$$\Sigma' = JW\Sigma W^\top J^\top, \quad (2)$$

where  $W$  is the view transformation, and  $J$  is the Jacobian of the projective transformation. Colors are computed via  $\alpha$ -blending to combine  $N$  Gaussians as:

$$C(u) = \sum_{i=1}^N c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad (3)$$

where  $c_i$  is the color defined by spherical harmonics, and  $\alpha_i$  is the multiplication of opacity and the transformed 2D Gaussian. Despite its effectiveness in scene reconstruction

and novel view synthesis, 3DGS struggles with real-world conditions such as motion blur or low-light conditions.

**Event Camera Model:** Each event  $e_k = (u_k, t_k, p_k)$  captures the pixel coordinate  $u_k$ , timestamp  $t_k$ , and polarity  $p_k$ , reflecting a brightness change exceeding the contrast threshold  $C$ . The brightness change for each event is given by:

$$\Delta L = L(u_k, t_k) - L(u_k, t_{k-1}) = p_k \cdot C, \quad (4)$$

where  $t_{k-1}$  represents the timestamp of the previous event at the same pixel. The instantaneous intensity image at time  $t$ , denoted as  $I(t) \in \mathbb{R}^{W \times H}$ , undergoes a logarithmic transformation to better represent perceptual brightness changes:

$$L(t) = \frac{\log I(t)}{g}, \quad (5)$$

where  $g = 2.2$  is a gamma correction factor applied in all experiments to linearize intensity values. This correction compensates for non-linear intensity encoding intended for display on sRGB monitors, with  $g = 2.2$  providing an effective approximation for standard gamma correction [30].

#### B. Temporal-to-Intensity Mapping of Exposure Events

Due to the limited texture information provided by motion events, we propose capturing exposure events during scene reconstruction by controlling the camera's aperture. High-quality grayscale images are then generated through temporal-to-intensity mapping of these exposure events,

achieved by dynamically adjusting the event camera's transmittance. Specifically, we introduce a Transmittance Adjustment (TA) device, where the transmittance rate  $TR(t)$  changes from 0 to 1 according to the function:

$$TR(t) = \begin{cases} f(t), & 0 \leq t \leq t_{\text{end}} \\ 1, & t > t_{\text{end}} \end{cases} \quad (6)$$

Each pixel triggers an event at a specific time  $t^*(u)$ , forming a temporal matrix. The photocurrent  $I_{\max}(u)$ , which represents the scene's intensity, can be derived from the event time  $t^*(u)$  as follows:

$$I_{\max}(x, y) = \frac{\exp((V_{\text{ref}} + V_{\text{thd}}) \cdot C_{\text{PD}}) - 1}{h(t^*(u))}, \quad (7)$$

where

$$h(t^*) = \int_0^{t^*(x,y)} TR(t) dt. \quad (8)$$

After mapping the time information to grayscale values, normalization is performed by adjusting  $I_{\max}(x, y)$  to the range  $[0, 1]$ , yielding a high-resolution grayscale image with adaptive dynamic range. The entire process is robust to various forms of degradation, including noise, contrast threshold variability, and pixel anomalies. The resulting temporal-to-intensity mapping of exposure events can further introduce additional constraints to optimize the event-based 3DGS, as detailed in Sec. III-C.

### C. Loss Functions

The loss function guiding the training process consists of two primary components: the motion event loss and the exposure event loss. The motion event loss is crucial for ensuring that the predicted brightness variations align with the motion events captured by the event camera. It is defined as:

$$L_{\text{evs,norm}} = \left\| \frac{\Delta \hat{L}(u_k)}{\|\Delta \hat{L}(u_k)\|_2} - \frac{\Delta L(u_k)}{\|\Delta L(u_k)\|_2} \right\|_2^2, \quad (9)$$

where  $\Delta \hat{L}(u_k)$  represents the predicted logarithmic brightness change, calculated as  $\Delta \hat{L}_k = \hat{L}(x_k, t_k) - \hat{L}(x_k, t_{k-1})$ . The predicted brightness values,  $\hat{L}(u_k, t_k)$  and  $\hat{L}(u_k, t_{k-1})$ , are obtained from the 3DGS model through volumetric rendering. This formulation ensures that the predicted brightness changes are consistent with the real-world motion event stream observed by the event camera, allowing for accurate motion capture under high-motion scenes.

In addition, the exposure event loss supervises the 3DGS reconstruction of high-quality frames obtained from a temporal-to-intensity mapping of exposure events. It is computed as the  $L_2$  loss (squared error) between the predicted image and the ground truth:

$$L_{\text{rgb}} = \frac{1}{N} \sum_{k=1}^N (I_{\text{pred}}(u_k) - I_{\text{gt}}(u_k))^2, \quad (10)$$

where  $I_{\text{pred}}(u_k)$  and  $I_{\text{gt}}(u_k)$  are the predicted and ground truth color values at pixel  $u_k$ , respectively, and  $N$  represents

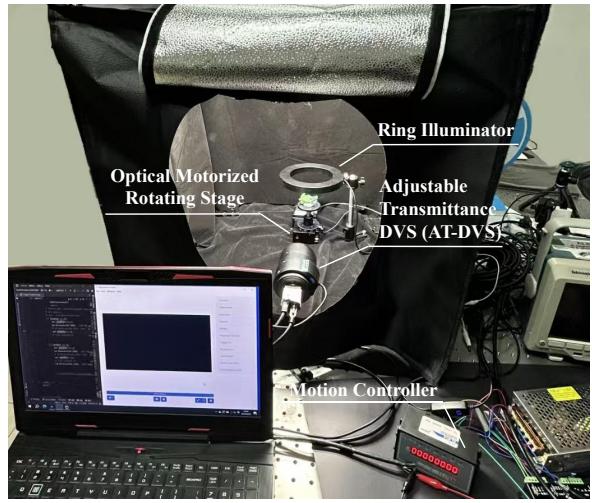


Fig. 2. Real-world data acquisition setup: The object is placed on a motorized optical rotation stage and illuminated by an overhead ring light to ensure uniform lighting. The scene is captured using an AT-DVS, implemented with a Prophesee Evaluation Kit 4 HD (EVK4) and an aperture shutter, facilitating high dynamic range event-based grayscale imaging through temporal-to-intensity mapping by exposure events.

the total number of pixels in the image. The final combined loss used during training is:

$$L = \lambda \cdot L_{\text{evs,norm}} + (1 - \lambda) \cdot L_{\text{rgb}}, \quad (11)$$

where  $\lambda = 0$  prioritizes high-quality texture reconstruction (High-Quality Reconstruction Mode),  $\lambda = 1$  focuses on fast motion capture (Fast Reconstruction Mode), and  $\lambda = 0.5$  balances both for a compromise between speed and quality (Balanced Hybrid Mode). By incorporating exposure events into the 3DGS framework, this loss adds a novel constraint that enhances texture reconstruction from sparse event data using a single event sensor.

### D. EME-3D: Exposure and Motion Events Dataset for 3D Reconstruction

We utilize an AT-DVS (Adjustable Transmittance DVS) to capture real-world sequences. Fig. 2 illustrates the experimental setup used for data acquisition. The optical motorized rotating stage in our setup offers a high precision of  $0.005^\circ$ . Before the experiment, the object is securely affixed to the stage using Blu-Tack to prevent any relative movement during rotation. Each object undergoes two recording phases, with the rotating stage reset to the zero position before each phase. In the first phase, the stage performs a full  $360^\circ$  rotation to capture motion events. During the second phase, the stage rotates in  $1.8^\circ$  increments, controlled by a motion controller, to capture exposure events at specific positions.

## IV. EXPERIMENTS

### A. Implementation Details

Our implementation is based on the official 3D-GS codebase [10]. For training in the fast reconstruction mode, which uses only motion event supervision, we run for 10,000 iterations with a 500-iteration warmup and an initial learning rate of  $1.6e^{-5}$ . For high-quality reconstruction mode and balanced hybrid mode, we run for 30,000 iterations with a

TABLE II  
QUALITATIVE COMPARISONS OF NEURAL RECONSTRUCTION.

| Metric    | Method                                        | Real-World EME-3D Dataset |              |              |              |              |              |              |              |              |              |
|-----------|-----------------------------------------------|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|           |                                               | car                       | motorcycle   | camera       | cat          | crab         | forklift     | racing car   | tortoise     | Average      |              |
| PSNR (dB) | E2VID [20], [31]+NeRF [9]                     | 22.57                     | 24.34        | 20.94        | 24.72        | 22.73        | 21.94        | 24.38        | 20.27        | 21.99        | 22.65        |
|           | EventNeRF [14]                                | 18.58                     | 23.50        | 22.66        | 17.97        | 19.48        | 19.01        | 19.82        | 15.82        | 18.98        | 19.54        |
|           | E2VID [20], [31]+3DGS [10]                    | 14.48                     | 17.79        | 18.01        | 14.58        | 14.82        | 14.51        | 17.13        | 16.12        | 15.80        | 15.92        |
|           | E2VID [20], [31]+SAM [32]+3DGS [10]           | 22.89                     | 25.22        | 21.20        | 25.14        | 22.97        | 21.95        | 24.97        | 20.89        | 22.08        | 23.03        |
|           | Ours (Fast Reconstruction Mode)               | 23.62                     | 26.99        | 25.69        | 26.81        | 25.26        | 23.08        | 25.91        | 24.57        | 25.04        | 25.22        |
| SSIM ↑    | <b>Ours (High-Qualiy Reconstruction Mode)</b> | <b>33.16</b>              | <b>35.89</b> | <b>35.11</b> | <b>31.92</b> | <b>33.58</b> | <b>32.60</b> | <b>34.82</b> | <b>35.54</b> | <b>32.66</b> | <b>33.92</b> |
|           | E2VID [20], [31]+NeRF [9]                     | 0.89                      | 0.90         | 0.84         | 0.91         | 0.89         | 0.78         | 0.88         | 0.87         | 0.83         | 0.87         |
|           | EventNeRF [14]                                | 0.56                      | 0.48         | 0.62         | 0.66         | 0.66         | 0.51         | 0.53         | 0.21         | 0.45         | 0.52         |
|           | E2VID [20], [31]+3DGS [10]                    | 0.12                      | 0.08         | 0.13         | 0.06         | 0.10         | 0.17         | 0.09         | 0.07         | 0.13         | 0.11         |
|           | E2VID [20], [31]+SAM [32]+3DGS [10]           | 0.92                      | 0.92         | 0.89         | 0.92         | 0.91         | 0.80         | 0.91         | 0.91         | 0.85         | 0.89         |
|           | <b>Ours (High-Qualiy Reconstruction Mode)</b> | <b>0.97</b>               | <b>0.98</b>  | <b>0.97</b>  | <b>0.97</b>  | <b>0.96</b>  | <b>0.93</b>  | <b>0.97</b>  | <b>0.98</b>  | <b>0.96</b>  | <b>0.97</b>  |
| LPIPS ↓   | E2VID [20], [31]+NeRF [9]                     | 0.12                      | 0.10         | 0.17         | 0.10         | 0.11         | 0.21         | 0.13         | 0.12         | 0.15         | 0.13         |
|           | EventNeRF [14]                                | 0.40                      | 0.30         | 0.34         | 0.34         | 0.33         | 0.45         | 0.53         | 0.83         | 0.49         | 0.45         |
|           | E2VID [20], [31]+3DGS [10]                    | 0.63                      | 0.63         | 0.60         | 0.64         | 0.62         | 0.60         | 0.61         | 0.62         | 0.57         | 0.61         |
|           | E2VID [20], [31]+SAM [32]+3DGS [10]           | 0.12                      | 0.09         | 0.14         | 0.10         | 0.11         | 0.22         | 0.10         | 0.09         | 0.15         | 0.12         |
|           | Ours (Fast Reconstruction Mode)               | 0.14                      | 0.10         | 0.15         | 0.10         | 0.12         | 0.24         | 0.11         | 0.11         | 0.20         | 0.14         |
|           | <b>Ours (High-Qualiy Reconstruction Mode)</b> | <b>0.10</b>               | <b>0.06</b>  | <b>0.07</b>  | <b>0.08</b>  | <b>0.09</b>  | <b>0.15</b>  | <b>0.06</b>  | <b>0.06</b>  | <b>0.10</b>  | <b>0.09</b>  |

500-iteration warmup and an initial learning rate of  $1.6e^{-4}$ . All experiments were performed on a single NVIDIA RTX 3090-Ti GPU.

### B. Baselines

We benchmark our approach against a NeRF-based method, EventNeRF [14], and a naive baseline, E2VID [20] + NeRF [9], which cascades the event-to-video pipeline E2VID into NeRF. Additionally, we compare a method that cascades E2VID with a vanilla 3D Gaussian Splatting approach [10]. We observed significant background noise in the images reconstructed by the learning-based E2VID, which adversely affects the accuracy of E2VID + 3DGS. To address this, we have introduced a stronger baseline by introducing the Segment Anything Model (SAM) [32] to segment and remove background noise from the grayscale images reconstructed by E2VID. The segmentation quality was manually double-checked and refined, forming our strongest baseline: E2VID + SAM + 3DGS. For real-world scenes, we retrained EventNeRF and all baselines on the EME-3D dataset, rendering RGB and depth images for comparison to ensure fairness and reliability.

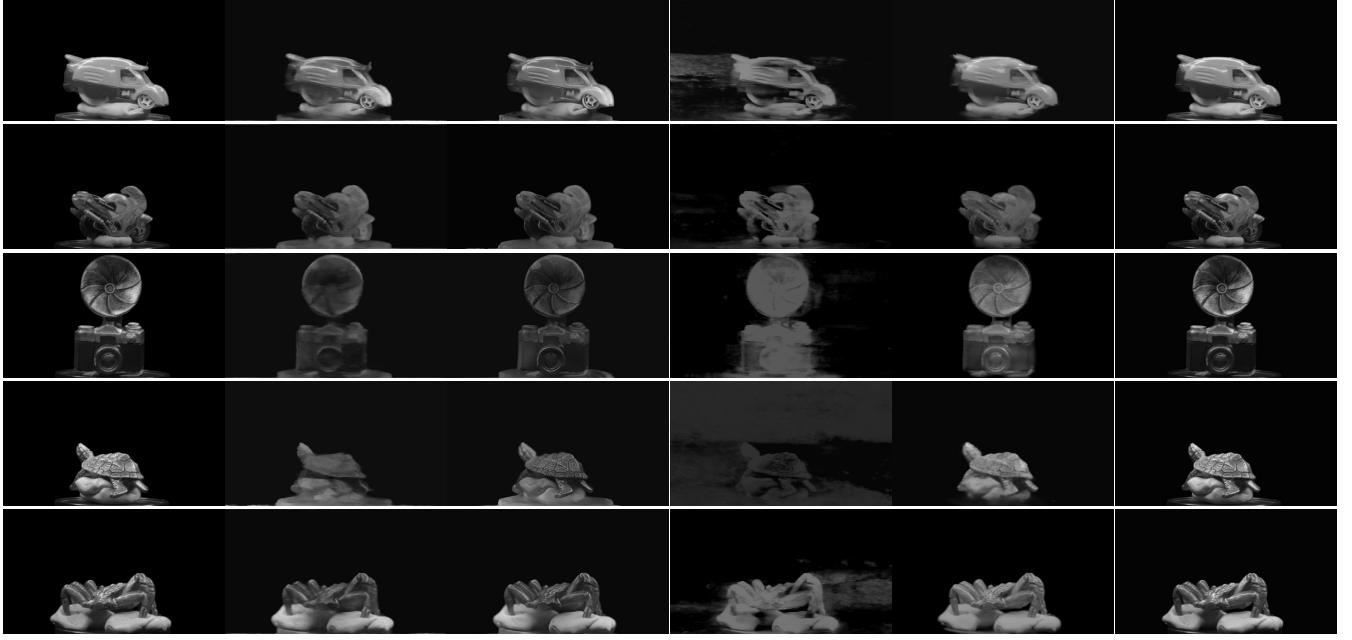
### C. Quantitative Comparison

As shown in Tab. II, we compare the proposed E-3DGS method with other baselines (see Sec. IV-B) on the real-world EME-3D dataset. In fast reconstruction mode, E-3DGS outperforms all baselines across all sequences in terms of reconstruction quality. On average, E-3DGS (Fast Reconstruction Mode), using only motion events, achieves a PSNR improvement of  $2.19dB$  ( $25.22$  vs.  $23.03$ ) over the strongest baseline E2VID + SAM + 3DGS, which utilizes additional learning-based events-to-video [20] and vision foundation model [32] techniques. Moreover, compared to the classic event-only method EventNeRF [14], E-3DGS (Fast Reconstruction Mode) shows a PSNR improvement of  $5.68dB$  ( $25.22$  vs.  $19.54$ ). In high-quality reconstruction mode, E-3DGS fully leverages the high-resolution spatial cues from exposure events to optimize the Gaussian ellipsoids and

distributions, resulting in a substantial further improvement in reconstruction quality. On average, E-3DGS achieves a PSNR increase of  $10.89dB$  ( $33.92$  vs.  $23.03$ ), an SSIM improvement of  $0.08$  ( $0.97$  vs.  $0.89$ ), and an LPIPS reduction of  $0.03$  ( $0.09$  vs.  $0.12$ ) compared to the strongest E2VID + SAM + 3DGS baseline. Notably, the training process of E-3DGS does not rely on the complex computations required by the learning-based E2VID or SAM, and E-3DGS also significantly outperforms the EventNeRF family in rendering speed ( $79.37$  FPS vs.  $0.03$  FPS).

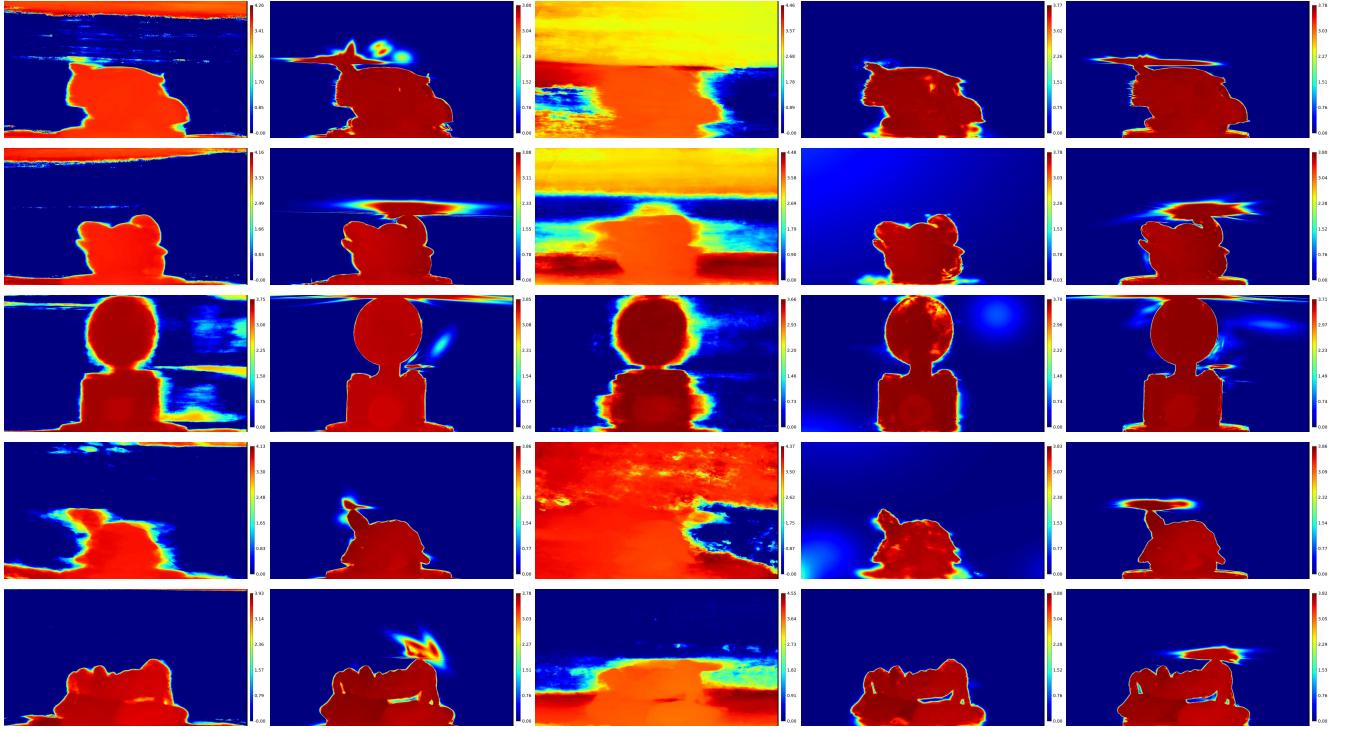
### D. Qualitative Comparison

We qualitatively compare the appearance and geometric reconstruction of five sequences from the real-world EME-3D dataset by visualizing the rendered results and depth maps. Fig. 3 shows that our E-3DGS, in both fast and high-quality reconstruction modes, preserves sharper, more consistent structures and cleaner backgrounds compared to EventNeRF [14]. When using motion events, E-3DGS avoids artifacts seen in EventNeRF’s rendering, such as with the car ( $1^{st}$  row) and the tortoise ( $4^{th}$  row). Our results also exhibit higher contrast and sharper spatial details, particularly in regions with highlights and reflections, like the camera ( $3^{rd}$  row) and the tortoise ( $4^{th}$  row). Compared to E2VID [20] + NeRF, E-3DGS produces clearer local details, such as the tortoise’s shell texture, which is completely lost in the  $4^{th}$  row of E2VID + NeRF. Furthermore, compared to E2VID + 3DGS, E-3DGS integrates exposure event information and reconstructs overall colors and brightness more accurately. For instance, in the  $3^{rd}$  row, E2VID + 3DGS shows white artifacts on the black camera body, while E-3DGS renders a clean, pure black camera body. As shown in Fig. 4, E-3DGS achieves cleaner backgrounds, fewer artifacts, and richer spatial details in geometric reconstruction compared to both EventNeRF and EventNeRF integrated with E2VID. EventNeRF’s depth maps struggle to distinguish foreground from background and suffer from noticeable noise, such as with the car ( $1^{st}$  row), motorcycle ( $2^{nd}$  row), tortoise ( $4^{th}$  row), and crab ( $5^{th}$  row). While the depth map quality of



**GT      E2VID+NeRF    E2VID+3DGS    EventNeRF    Ours (Fast)    Ours (HQ)**

Fig. 3. Qualitative comparison of appearance reconstruction on the real-world EME-3D dataset. The E2VID + NeRF method successfully reconstructs the overall scene but lacks fine local details, such as the tortoise’s shell texture. EventNeRF exhibits noticeable artifacts in both the car and the tortoise. In contrast, our proposed E-3DGS, in both fast and high-quality reconstruction modes, preserves sharper and more consistent structures, while maintaining cleaner backgrounds. Moreover, compared to E2VID + 3DGS, E-3DGS leveraging exposure events demonstrates superior detail and brightness reconstruction.



**E2VID+NeRF      E2VID+3DGS      EventNeRF      Ours (Fast)      Ours (HQ)**

Fig. 4. Qualitative comparison of geometry reconstruction on the real-world EME-3D dataset. Both E2VID+NeRF and EventNeRF, which are based on NeRF, struggle to separate the foreground from the background and are affected by noticeable noise. In contrast, the physics-based 3DGS method handles geometry reconstruction more effectively. Compared to E2VID+3DGS, our E-3DGS excels in capturing high-frequency spatial details, such as the crab’s body and the texture of the base.

E2VID + 3DGS is comparable to E-3DGS, E-3DGS excels in high-frequency spatial details. For example, in the crab sample, E-3DGS successfully reconstructs the gap between the crab’s body and the base, while E2VID + 3DGS confuses the subject with the background.

### E. Ablations

In the ablation study, unless otherwise specified, we train on each sequence of the real-world EME-3D dataset and report the average accuracy across all sequences. We investigate various aspects, including point cloud initialization

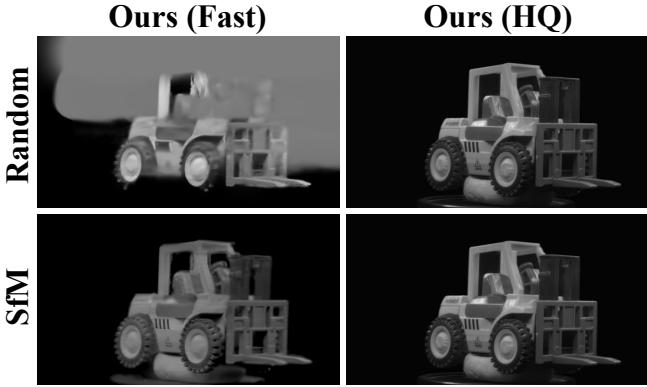


Fig. 5. Qualitative comparison of different initialization methods.

TABLE III

IMPACT OF INITIALIZATION METHODS ON FAST AND HQ MODES.

| Metric    | Method      | Random | SfM [12] | Difference |
|-----------|-------------|--------|----------|------------|
| PSNR (dB) | Ours (Fast) | 15.72  | 25.22    | + 60 %     |
|           | Ours (HQ)   | 33.83  | 33.92    | + 0.3%     |
| SSIM ↑    | Ours (Fast) | 0.21   | 0.89     | + 324 %    |
|           | Ours (HQ)   | 0.96   | 0.97     | + 1.04 %   |
| LPIPS ↓   | Ours (Fast) | 0.58   | 0.14     | 75.9 %     |
|           | Ours (HQ)   | 0.09   | 0.09     | -          |

methods for event-based 3DGS, the accuracy variation in the balanced hybrid mode of E-3DGS, and the accuracy difference between novel view synthesis and 3D reconstruction tasks using E-3DGS.

**Point Cloud Initialization.** In Tab. III, we explore the impact of different point cloud initialization methods on rendering performance in event-based 3DGS. Compared to random initialization, using a sparse point cloud from SfM [12] significantly improves the rendering accuracy when using motion events, with PSNR elevating to 25.22dB from 15.72dB. However, in high-quality reconstruction mode, the PSNR is not sensitive to whether the initialization is from SfM or random point clouds (33.82dB vs. 33.92dB). As shown in Fig. 5, we visualize the effects of different initialization methods on event-based 3DGS rendering. When using random initialization, motion event-based 3DGS produces noticeable artifacts and limited details (top-left vs. bottom-left). In contrast, after integrating the spatially rich exposure events, E-3DGS becomes more robust to the initialization method, with no significant differences in rendering results (top-right vs. bottom-right). To further illustrate the advantages of incorporating exposure events in 3D reconstruction, we compare NeRF-based and event-to-video-based point cloud initialization methods in Tab. IV. Initializing the point cloud using EventNeRF-rendered results, compared to random initialization, provides a modest improvement in PSNR (17.13dB vs. 15.72dB), though it requires additional training for EventNeRF with limited gains. Using E2VID [20] to convert events into images followed by SfM for point cloud initialization further enhances accuracy (23.47dB vs. 15.72dB), but this process also involves additional learning-based computation. By fully leveraging

TABLE IV  
ABLATION STUDY ON DIFFERENT INITIALIZATION METHOD.

| Method                | PSNR ↑       | SSIM ↑      | LPIPS ↓     |
|-----------------------|--------------|-------------|-------------|
| Random Initialization | 15.72        | 0.21        | 0.58        |
| EventNeRF [14]        | 17.13        | 0.72        | 0.29        |
| E2VID+SfM [12]        | 23.47        | 0.87        | 0.15        |
| Exposure+SfM [12]     | <b>25.22</b> | <b>0.89</b> | <b>0.14</b> |

Effects of integrating motion events with exposure events on 3D reconstruction

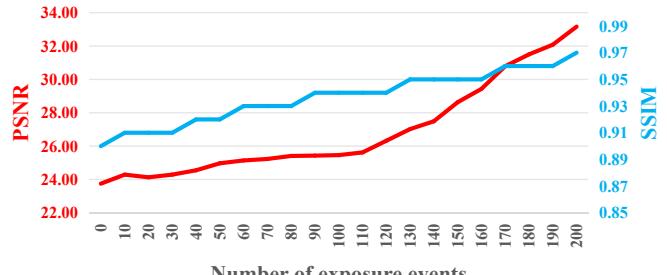


Fig. 6. Effects of integrating motion events with exposure events on 3D reconstruction for the real-world car sequence.

TABLE V

QUANTITATIVE RESULTS OF NOVEL VIEW RECONSTRUCTION.

| Sequence   | Ours (Fast Reconstruction) |       |       |       | Ours (High-Quality Reconstruction) |       |       |       |
|------------|----------------------------|-------|-------|-------|------------------------------------|-------|-------|-------|
|            | PSNR                       |       | SSIM  |       | PSNR                               |       | SSIM  |       |
|            | Given                      | Novel | Given | Novel | Given                              | Novel | Given | Novel |
| car        | 22.98                      | 22.96 | 0.88  | 0.87  | 25.60                              | 25.42 | 0.91  | 0.91  |
| motorcycle | 25.22                      | 24.94 | 0.90  | 0.89  | 27.34                              | 27.06 | 0.92  | 0.92  |
| camera     | 23.99                      | 24.05 | 0.87  | 0.87  | 26.42                              | 26.43 | 0.90  | 0.90  |
| cat        | 25.35                      | 25.34 | 0.91  | 0.90  | 26.75                              | 27.34 | 0.93  | 0.92  |
| crab       | 22.21                      | 22.19 | 0.87  | 0.87  | 23.11                              | 23.05 | 0.89  | 0.89  |
| forklift   | 19.92                      | 19.90 | 0.76  | 0.76  | 21.37                              | 21.23 | 0.81  | 0.81  |
| racing car | 23.83                      | 23.74 | 0.88  | 0.88  | 25.73                              | 25.57 | 0.91  | 0.91  |
| tortoise   | 22.55                      | 22.44 | 0.88  | 0.88  | 24.26                              | 24.10 | 0.91  | 0.91  |
| train      | 21.80                      | 21.77 | 0.82  | 0.82  | 23.41                              | 23.30 | 0.85  | 0.85  |
| Average    | 23.09                      | 23.04 | 0.86  | 0.86  | 24.89                              | 24.83 | 0.89  | 0.89  |

the rich spatial cues embedded in exposure events, the proposed exposure event temporal-spatial mapping combined with SfM initialization yields the greatest improvement in reconstruction quality (25.22dB vs. 15.72dB). This certifies the critical role of exposure events in enhancing the quality of explicit 3D reconstruction from events.

**Balanced Hybrid Mode.** We further look into how to balance reconstruction quality and the number of exposure event frames in E-3DGS. To handle fast and complex scenes, E-3DGS includes a balanced hybrid mode. In this mode, we control the aperture to capture a small number of exposure events during the initial phase of a sample, followed by collecting only motion events to achieve high-quality reconstruction of fast-moving scenes. In Fig. 6, we show the impact of the number of exposure event frames on rendering quality. As the number of exposure event frames increases, more accurate spatial information is used to constrain the optimization of the Gaussian splatting, leading to a steady improvement in PSNR and SSIM of the rendered images. In other words, introducing even a small number of exposure event frames provides a positive gain in 3D reconstruction quality compared to using none. This highlights the effectiveness and potential of the balanced hybrid mode for handling fast-motion scenes.

**Novel View Synthesis.** In this task, we trained the model using half of the odd-numbered frames (Given) and tested its ability to generate novel views using the remaining even-numbered frames (Novel). As shown in Tab. V, E-3DGS has excellent novel view synthesis performance in both fast reconstruction mode and high-quality reconstruction mode. On average, E-3DGS using motion events exhibits only a  $0.05dB$  drop in PSNR for novel view synthesis compared to the given frames ( $23.04dB$  vs.  $23.09dB$ ). Similarly, E-3DGS using exposure events shows a  $0.06dB$  PSNR drop for novel views compared to the given frames ( $24.83dB$  vs.  $24.89dB$ ). This demonstrates E-3DGS’s strong capabilities in both 3D reconstruction and novel view synthesis tasks.

## V. CONCLUSION

In this work, we have presented E-3DGS, the first method to integrate 3DGS with exposure events for explicit scene reconstruction using a single pure high-resolution event sensor. By partitioning events into motion and exposure categories, E-3DGS leverages motion events for efficient 3D Gaussian Splatting (3DGS) reconstruction and utilizes exposure events to enhance reconstruction quality through high-resolution temporal-to-intensity mapping. Our framework is versatile, capable of operating on motion events alone or adopting a hybrid mode that balances reconstruction quality and speed by combining initial exposure events with high-speed motion events. Additionally, we introduced EME-3D, a real-world dataset with exposure and motion events, camera calibration parameters, and sparse point clouds to support further research in this domain. E-3DGS outperforms event-based NeRF in both reconstruction speed and quality while being more cost-effective than methods that require both event and RGB data. By fully utilizing both motion and exposure events, E-3DGS sets a new benchmark for event-based 3D reconstruction, demonstrating robust performance in challenging conditions with reduced hardware demands. In the future, we intend to explore the application and enhancement of E-3DGS based on exposure events for dense robotic perception, aiming to achieve higher-quality large-scale dense 3D reconstruction using purely event-based data.

## REFERENCES

- [1] A. Rosinol, J. J. Leonard, and L. Carlone, “NeRF-SLAM: Real-time dense monocular SLAM with neural radiance fields,” in *Proc. IROS*, 2023, pp. 3437–3444.
- [2] Z. Zhu *et al.*, “NICE-SLAM: Neural implicit scalable encoding for SLAM,” in *Proc. CVPR*, 2022, pp. 12776–12786.
- [3] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, “iNeRF: Inverting neural radiance fields for pose estimation,” in *Proc. IROS*, 2021, pp. 1323–1330.
- [4] L. Xu *et al.*, “VR-NeRF: High-fidelity virtualized walkable spaces,” in *Proc. SIGGRAPH Asia* 2023, 2023, pp. 1–12.
- [5] J. J. LaViola Jr, “Bringing VR and spatial 3D interaction to the masses through video games,” *IEEE Computer Graphics and Applications*, vol. 28, no. 5, pp. 10–15, 2008.
- [6] J. Kerr, C. M. Kim, K. Goldberg, A. Kanazawa, and M. Tancik, “LERF: Language embedded radiance fields,” in *Proc. ICCV*, 2023, pp. 19672–19682.
- [7] K. Liu *et al.*, “Weakly supervised 3D open-vocabulary segmentation,” in *Proc. NeurIPS*, vol. 36, 2023, pp. 53433–53456.
- [8] M. Liu *et al.*, “OpenShape: Scaling up 3D shape representation towards open-world understanding,” in *Proc. NeurIPS*, vol. 36, 2024.
- [9] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [10] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3D gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, 2023.
- [11] S. Klenk, L. Koestler, D. Scaramuzza, and D. Cremers, “E-NeRF: Neural radiance fields from a moving event camera,” *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1587–1594, 2023.
- [12] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: exploring photo collections in 3D,” *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 835–846, 2006.
- [13] G. Gallego *et al.*, “Event-based vision: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.
- [14] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik, “EventNeRF: Neural radiance fields from a single colour event camera,” in *Proc. CVPR*, 2023, pp. 4992–5002.
- [15] I. Hwang, J. Kim, and Y. M. Kim, “Ev-NeRF: Event based neural radiance field,” in *Proc. WACV*, 2023, pp. 837–847.
- [16] Y. Qi, L. Zhu, Y. Zhang, and J. Li, “E<sup>2</sup>NeRF: Event enhanced neural radiance fields from blurry images,” in *Proc. ICCV*, 2023, pp. 13208–13218.
- [17] T. Xiong *et al.*, “Event3DGS: Event-based 3D gaussian splatting for high-speed robot egomotion,” in *Proc. CoRL*, 2024.
- [18] J. Wang, J. He, Z. Zhang, M. Sun, J. Sun, and R. Xu, “EvGGS: A collaborative learning framework for event-based generalizable gaussian splatting,” *arXiv preprint arXiv:2405.14959*, 2024.
- [19] W. Yu, C. Feng, J. Tang, X. Jia, L. Yuan, and Y. Tian, “EvaGaussians: Event stream assisted gaussian splatting from blurry images,” *arXiv preprint arXiv:2405.20224*, 2024.
- [20] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, “Events-to-video: Bringing modern computer vision to event cameras,” in *Proc. CVPR*, 2019, pp. 3857–3866.
- [21] J. Hidalgo-Carrió, G. Gallego, and D. Scaramuzza, “Event-aided direct sparse odometry,” in *Proc. CVPR*, 2022, pp. 5771–5780.
- [22] P. Achlioptas, O. Diamanti, I. Mitiagkas, and L. Guibas, “Learning representations and generative models for 3D point clouds,” in *Proc. ICML*, 2018, pp. 40–49.
- [23] S. Liu, T. Li, W. Chen, and H. Li, “A general differentiable mesh renderer for image-based 3D reasoning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 50–62, 2022.
- [24] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh, “Neural volumes: Learning dynamic renderable volumes from images,” *arXiv preprint arXiv:1906.07751*, 2019.
- [25] V. Sitzmann, J. Thies, F. Heide, M. Nießner, G. Wetzstein, and M. Zollhofer, “DeepVoxels: Learning persistent 3D feature embeddings,” in *Proc. CVPR*, 2019, pp. 2437–2446.
- [26] J. Wu, S. Zhu, C. Wang, and E. Y. Lam, “Ev-GS: Event-based gaussian splatting for efficient and accurate radiance field rendering,” *arXiv preprint arXiv:2407.11343*, 2024.
- [27] H. Deguchi, M. Masuda, T. Nakabayashi, and H. Saito, “E2GS: Event enhanced gaussian splatting,” in *Proc. ICIP*, 2024, pp. 1676–1682.
- [28] Y. Weng, Z. Shen, R. Chen, Q. Wang, and J. Wang, “EaDeblur-GS: Event assisted 3D deblur reconstruction with gaussian splatting,” *arXiv preprint arXiv:2407.13520*, 2024.
- [29] Y. Bao, L. Sun, Y. Ma, and K. Wang, “Temporal-mapping photography for event cameras,” in *Proc. ECCV*, 2024.
- [30] International Standard IEC 61966-2-1:1999: Amendment 1 - Multimedia systems and equipment – Colour measurement and management – Part 2-1: Colour management – Default RGB colour space – sRGB, International Electrotechnical Commission Std. TC 100/TA 2, 2003.
- [31] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, “High speed and high dynamic range video with an event camera,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1964–1980, 2021.
- [32] Y. Xiong *et al.*, “EfficientSAM: Leveraged masked image pretraining for efficient segment anything,” in *Proc. CVPR*, 2024, pp. 16111–16121.