Exploiting Event-Based Geometry Cues for Deblurring 3D Gaussian Splatting

Gu Gong, Fangyuan Wang, Zhen He, Qiang Wang, *Member, IEEE*
and David Navarro-Alarcon, *Senior Member, IEEE*

***Abstract*—3D Gaussian Splatting (3DGS) achieves real-time novel view synthesis but relies on Structure-from-Motion (SfM) for camera poses, which fails under severe motion blur. Recent event-aided approaches offer blur-free observations but do not fully exploit the geometric information embedded in event streams, limiting reconstruction quality. In this paper, we propose EEDGS, a framework that exploits event-based geometry cues for robust 3DGS reconstruction under severe motion blur. We first extract geometry cues from event streams through a feed-forward vision transformer to directly infer camera poses and point clouds, completely bypassing blur-corrupted RGB-based SfM. Building upon these event-derived geometric priors, we further develop a differentiable trajectory optimization module that jointly refines camera poses with 3D Gaussians in an alternating manner. Extensive experiments on synthetic and real-world datasets demonstrate that EEDGS significantly outperforms state-of-the-art methods, improving PSNR by up to 5.15 dB and reducing absolute trajectory error by over 58% compared to the best existing approaches under severe motion blur.**

***Index Terms*—Event Camera, 3D Gaussian Splatting, Motion Deblurring, Pose Refinement, Novel View Synthesis**

I. Introduction

RECENTLY, 3D Gaussian Splatting (3DGS) [1] has emerged as a powerful representation for novel view synthesis, achieving real-time rendering while maintaining high visual fidelity. By representing scenes as collections of anisotropic 3D Gaussians optimized through differentiable rasterization, 3DGS enables applications ranging from virtual reality to autonomous navigation. However, the quality of 3DGS reconstruction fundamentally relies heavily on two prerequisites, accurate camera poses and reliable initial point cloud, both typically obtained through Structure-from-Motion (SfM) methods like COLMAP [2]. When input images suffer from motion blur—a common occurrence in handheld capture or fast-moving platforms—SfM methods usually fail to extract precise camera poses and point cloud, leading to complete reconstruction failure as illustrated in Fig. 1.

To tackle this issue, recent deblurring 3DGS approaches [3], [4] explicitly model the physical blur process by optimizing camera trajectories to synthesize motion blur. While effective for mild blur, these methods face a critical "chicken-and-egg" dilemma: they rely heavily on accurate camera poses to produce promising deblurring images, but inferring accurate camera poses via standard SfM also depends on blur-free images. Consequently, under severe motion blur, these methods inevitably fail due to the collapse of camera pose initialization.

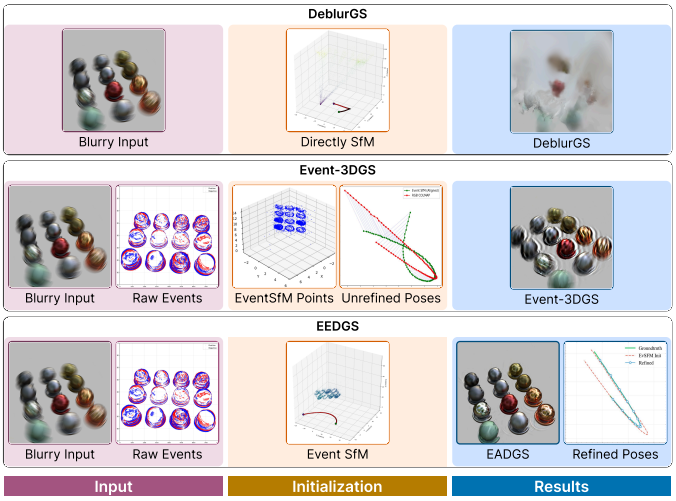

Fig. 1. Comparison of three paradigms for 3DGS reconstruction from motion-blurred inputs. Top: RGB-based methods suffer from catastrophic SfM failure under severe blur, yielding poor initialization and degraded reconstruction. Middle: existing event-aided methods use accumulated events for indirect initialization but keep camera poses fixed during training, resulting in limited reconstruction quality. Bottom: our EEDGS employs event-based SfM for robust initialization and jointly optimizes camera trajectories with scene geometry, achieving high-fidelity reconstruction.

Motivated by the success of event cameras in highly dynamic scenes [5], event-aided approaches have drawn increasing interest in a wide range of applications to exploit the complementary information provided by event streams. However, current state-of-the-art event-deblurring approaches [6]–[8] remain suboptimal. These methods typically adopt an "indirect" camera pose initialization strategy that first deblurs images using events and then performs standard SfM—a two-stage paradigm that is computationally expensive and prone to error accumulation from deblurring artifacts. More critically, the initialized poses are kept fixed during 3DGS optimization, preventing joint refinement of camera trajectories and scene geometry, which further limits reconstruction quality when the initial estimates are imperfect.

To address these limitations, we propose EEDGS (**E**vent-

Corresponding authors: David Navarro-Alarcon and Qiang Wang.
G. Gong is with the Dept. of Control Science and Engineering, Harbin Institute of Technology, China and the Dept. of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: gonggu@stu.hit.edu.cn).
F. Wang and D. Navarro-Alarcon are with the Dept. of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: fangyuan.wang@connect.polyu.hk, dnavar@polyu.edu.hk).
Q. Wang and Z. He are with the Dept. of Control Science and Engineering, Harbin Institute of Technology, China (e-mail: wangqiang@hit.edu.cn, hezhen@hit.edu.cn).

Enhanced **D**eblurring **G**aussian **S**platting), a unified framework that exploits event-based geometry cues for robust 3DGS reconstruction under severe motion blur. Our key insight is that event streams encode rich geometric information that can serve as reliable priors throughout the entire reconstruction pipeline. Concretely, we first extract geometry cues, including camera poses and scene point clouds, directly from event streams via a feed-forward vision transformer, completely bypassing blur-corrupted RGB-based SfM. Building upon these event-derived geometric priors, we further develop a differentiable trajectory optimization module that jointly refines camera poses with scene geometry during 3DGS training.

Our main contributions are summarized as follows:

- We propose EEDGS, a unified framework that exploits event-based geometry cues for 3DGS reconstruction under severe motion blur. We demonstrate that the geometric information embedded in event streams can serve as reliable priors for both scene initialization and subsequent trajectory optimization.
- We develop an event-based initialization pipeline that infers camera pose and point cloud directly from events, completely bypassing blur-corrupted RGB images.
- We introduce a pose refinement module that is compatible with the 3DGS framework to achieve joint optimization of scene geometry and camera trajectories.
- We conduct experiments on both synthetic and real-world datasets. The results demonstrate the effectiveness of our method in terms of both rendering quality and pose estimation accuracy.

II. Related Work

In this section, we first review event-based novel view synthesis methods. Then, we discuss recent advances for motion deblurring in neural rendering.

A. Event-Based Novel View Synthesis

Event cameras record per-pixel brightness changes asynchronously with microsecond-level temporal resolution, offering advantages including high dynamic range and inherent immunity to motion blur [9]. These properties have motivated a growing body of work on event-based novel view synthesis.

Early methods focused on neural radiance fields (NeRF). Ev-NeRF [10] first demonstrated that NeRF can be trained directly from raw event streams by rendering intensity differences between adjacent frames to match accumulated events. E-NeRF [11] extended this paradigm to moving cameras by incorporating an event generation model to provide supervision. EventNeRF [12] further demonstrated that a single color event camera suffices for NeRF training by accumulating events into dense representations, demonstrating robustness under fast motion and low-light conditions. To handle real-world degradations, Robust e-NeRF [13] introduced a more realistic event generation model that directly reconstructs NeRF from sparse and noisy events under non-uniform motion.

Inspired by the great success of 3D Gaussian Splatting (3DGS), event-based 3DGS methods have attracted increasing interests. Event-3DGS [14] pioneered this direction by modeling the relationship between Gaussian primitives and event streams, enabling efficient optimization through manually derived gradients. EaDeblur-GS [8] incorporated events to assist deblurring in 3DGS reconstruction, using an adaptive deviation estimator to model camera shake during exposure. EvaGaussians [15] proposed an event stream assisted training strategy for 3DGS to recover sharp novel views from blurry images. EventSplat [16] further advanced the field by achieving real-time novel view synthesis from moving event cameras. However, these methods commonly rely on accurate camera poses from external sensors or pre-calibrated setups, suffering severe performance drop under inaccurate poses.

B. Motion Deblurring in Neural Rendering

Motion blur arises from temporal integration during camera exposure and severely degrades reconstruction quality in neural rendering. Deblur-NeRF [17] first addressed this problem by modeling the physical blur formation process and jointly optimizing a blur kernel with the radiance field. Subsequent works extended this paradigm to 3D Gaussian Splatting. Specifically, BAD-Gaussians [4] performed bundle adjustment jointly with Gaussian optimization, while Deblurring 3DGS [3] modeled camera trajectories as continuous splines to synthesize blur during training. Despite promising performance, these methods share a critical dependency on SfM for 3DGS initialization. To alleviate this reliance, BARF [18] and NoPe-NeRF [19] achieved joint optimization of camera poses with neural radiance fields, while CF-3DGS [20], LocalRF [21], and Gaussian Splatting SLAM [22] extended SfM-free reconstruction to 3DGS. Feed-forward methods such as DUSt3R [23] and MASt3R [24] further demonstrated that dense 3D reconstruction and pose estimation can be achieved without iterative SfM. However, these methods fundamentally depend on RGB-based features and thus degrade under severe motion blur.

Recently, event-aided approaches attempt to break this dependency. Particularly, E2NeRF [6] used event streams to enhance blurry images before feeding them to NeRF, while EvDeblurNeRF [7] incorporated event streams as supervision during training. Nevertheless, these methods still rely on intermediate deblurring techniques (*e.g.*, Event Double Integral [25]) to reconstruct sharp images for SfM initialization. This two-stage paradigm suffers two major limitations. First, original blurry images may lack multi-view consistency, producing erroneous camera poses. Second, the estimated poses remain fixed during 3DGS optimization. As a result, pose errors are accumulated and even amplified during training, producing blurry reconstruction results.

III. Preliminaries

This section introduces the foundational concepts underlying our framework. We first describe the event camera model that provides blur-free observations (Sec. III-A), then formulate the motion blur formation process that our method aims to invert (Sec. III-B), and finally review 3D Gaussian Splatting as the scene representation we build upon (Sec. III-C).

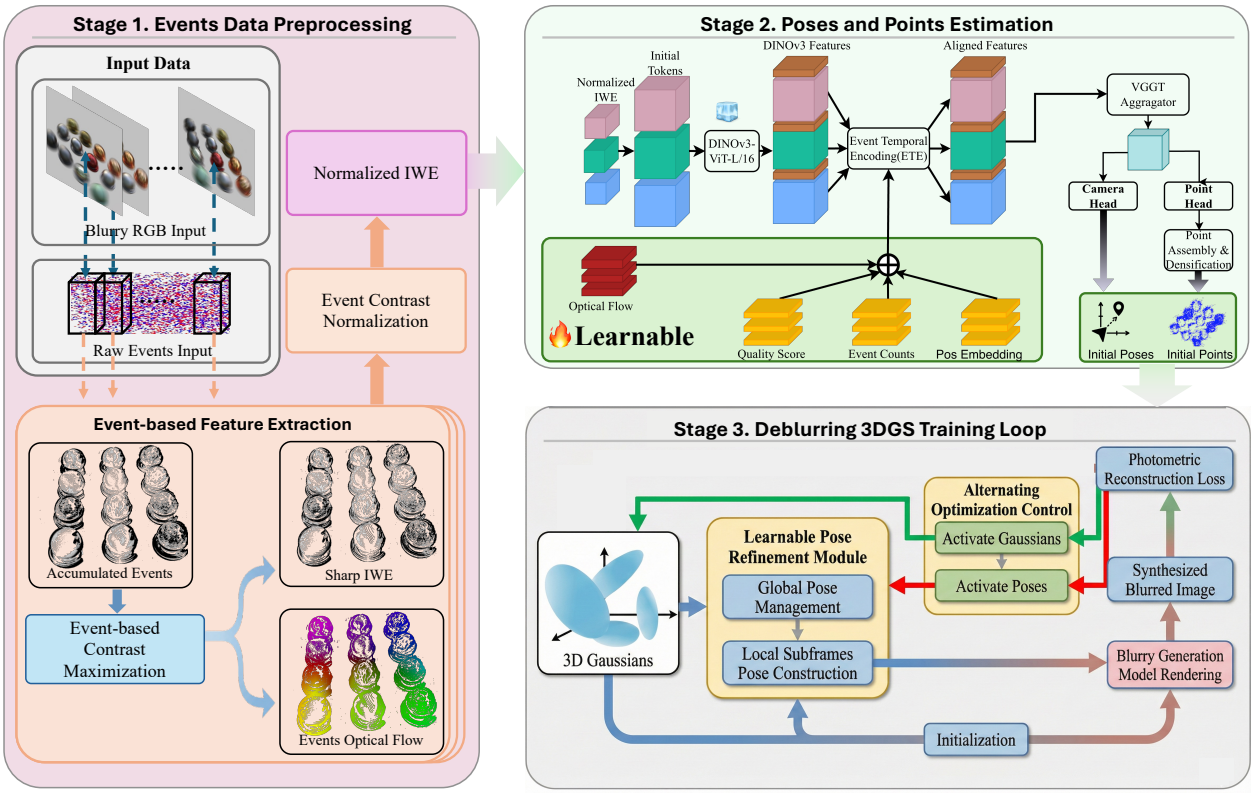

Fig. 2. Overview of the proposed **EEDGS**. The framework consists of three stages: Stage 1 preprocesses raw events via contrast maximization and normalization to obtain sharp event images and dense optical flow fields. Stage 2 employs a feed-forward vision transformer to predict initial camera pose and point cloud from the event descriptor produced by Stage 1. Stage 3 employs the results from Stage 2 for initialization and then jointly optimizes 3D Gaussians with camera trajectories via a learnable pose refinement.

A. Event Camera Model

Event cameras asynchronously record per-pixel brightness changes with microsecond temporal resolution. An event $e_k = (x_k, y_k, t_k, p_k)$ is triggered at pixel (x_k, y_k) and time t_k when the logarithmic brightness change exceeds a threshold C:

$$L(x_k, y_k, t_k) - L(x_k, y_k, t_k - \delta t) = p_k C, \tag{1}$$

where $p_k \in \{-1, +1\}$ indicates the polarity of brightness change. Since event cameras respond to instantaneous brightness changes, they are inherently immune to motion blur.

B. Motion Blur Formation

Motion blur arises from the temporal integration of scene radiance during camera exposure. The observed blurry image $\mathbf{B}$ is formed by averaging instantaneous sharp images over the exposure interval:

$$\mathbf{B} \approx \frac{1}{N} \sum_{n=1}^{N} \mathbf{I}(\mathbf{T}_n), \tag{2}$$

where $\mathbf{I}(\mathbf{T}_n)$ denotes the sharp image rendered at subframe pose $\mathbf{T}_n$, and $\{\mathbf{T}_n\}_{n=1}^{N}$ are camera poses distributed along the motion trajectory during exposure.

C. 3D Gaussian Splatting

3D Gaussian Splatting [1] represents scenes using a set of 3D Gaussians $\mathcal{G} = \{G_k\}_{k=1}^{K}$. Each Gaussian is formulated using a 3D covariance matrix $\mathbf{\Sigma}$ centered at point $\boldsymbol{\mu}$:

$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \tag{3}$$

The covariance matrix can be decomposed as $\mathbf{\Sigma} = \mathbf{RSS}^\top\mathbf{R}^\top$, where $\mathbf{R}$ is a rotation matrix and $\mathbf{S}$ is a diagonal scaling matrix. In addition, each Gaussian stores an opacity value α and spherical harmonic (SH) coefficients for view-dependent color rendering.

For rendering, 3D Gaussians are projected onto the image plane. Given camera pose $\mathbf{T}$, the rendered color at pixel $\mathbf{r}$ is computed via differentiable alpha-blending:

$$\mathbf{c}(\mathbf{r}) = \sum_{i=1}^{K} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \tag{4}$$

where c_i and α_i are the color and opacity of the i-th Gaussian along the ray. The Gaussians are optimized by minimizing the photometric loss between rendered and ground truth images, with adaptive densification and pruning.

IV. Method

A. Overview

Given a set of blurry RGB images $\{\mathbf{B}_i\}_{i=1}^{M}$ and a synchronized event stream $\mathcal{E}$, our goal is to reconstruct a sharp

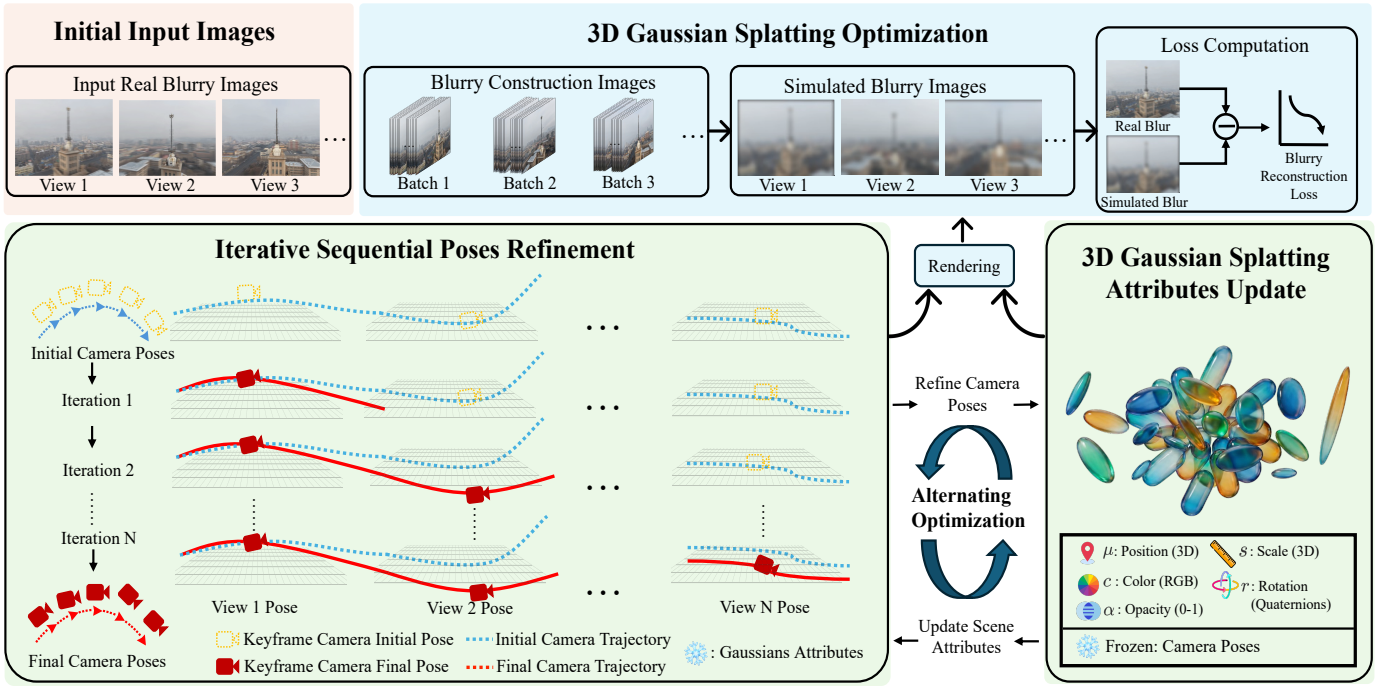

Fig. 3. Overview of the training process. During training, camera poses and Gaussian attributes are alternately optimized to minimize the loss.

3D Gaussian scene representation $\mathcal{G}$ along with accurate camera poses $\{\mathbf{T}_i\}_{i=1}^{M}$. As illustrated in Fig. 2, our method consists of three stages: (1) *Event Data Preprocessing*, which applies contrast maximization and event image normalization to extract sharp event images and dense optical flow fields from raw event streams; (2) *Camera Pose and Point Cloud Estimation*, which employs a feed-forward vision transformer to predict initial camera poses and dense point maps from the event images, followed by confidence-based filtering to construct the initial point cloud; and (3) *Deblurring 3DGS Training*, which jointly optimizes 3D Gaussians and camera trajectories through learnable pose refinement with alternating optimization.

B. Event Data Preprocessing

To bypass blur-corrupted RGB images, we extract geometric information—camera poses and 3D point clouds—directly from the event stream. However, leveraging event data for visual geometry estimation poses two fundamental challenges. **(i) Inherent sparsity:** Events are triggered only at edge locations where brightness changes exceed the threshold C, producing highly sparse and unstructured data. **(ii) Domain gap:** Although event data can be formulated as image-like representations, they exhibit fundamentally different distributions from natural images—zero-centered, variable dynamic range, and unbounded values—making models pretrained on natural images ineffective without explicit adaptation. To address these two challenges, we first employ contrast maximization to convert the sparse event stream into sharp, dense images (Sec. IV-B1), and then a learnable normalization layer is adopted bridge the discrepancy with the natural image domain for subsequent processes (Sec. IV-B2).

1) Contrast Maximization for Event Alignment: The input event stream $\mathcal{E}$ is first partitioned into temporal bins aligned with the exposure intervals of the blurry RGB frames. Within each bin, contrast maximization [26] jointly estimates the per-pixel optical flow $\mathbf{v}(x, y)$ and warps all events to a reference time t_{ref}:

$$\begin{pmatrix} x'_k \\ y'_k \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} + (t_{\text{ref}} - t_k) \cdot \mathbf{v}(x_k, y_k), \tag{5}$$

The warped events are then accumulated into a dense Image of Warped Events (IWE):

$$I^{\text{event}}(x, y) = \sum_k p_k \cdot \mathcal{K}(x - x'_k, y - y'_k), \tag{6}$$

where $\mathcal{K}(\cdot)$ is a bilinear interpolation kernel and the summation is calculated over all events in the bin. The optimal flow is obtained by maximizing the IWE variance, which corresponds to maximizing edge sharpness. This yields a sharp event image I_i^{event} and a dense optical flow field $\mathbf{v}_i$ for each temporal bin.

2) Event Image Normalization: While contrast maximization resolves the sparsity challenge by producing dense, sharp event images, the resultant IWE representations still suffer notable domain gap against natural images, which hinders the adoption of existing models pretrained on natural images. To bridge this gap, we develop a learnable normalization:

$$\hat{I}_i^{\text{event}} = \frac{I_i^{\text{event}} - \mu(I_i^{\text{event}})}{2\sigma(I_i^{\text{event}}) + \epsilon} \cdot \boldsymbol{\gamma}_e + \boldsymbol{\beta}_e, \tag{7}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation values, $\boldsymbol{\gamma}_e, \boldsymbol{\beta}_e \in \mathbb{R}^3$ are learnable parameters. The factor of 2 accounts for the wider dynamic range of event accumulations. Unlike batch normalization, this normalization operates per image to preserve the contrast structure of each event frame.

After event data processing, the output for each temporal bin i consists of a normalized event image $\hat{I}_i^{\text{event}}$ and an optical flow map $\mathbf{v}_i$, which serve as input for the subsequent stage.

C. Camera Pose and Point Cloud Estimation

Given the normalized event images $\{\hat{I}_i^{\text{event}}\}_{i=1}^{S}$ from the preprocessing stage, we employ VGGT [27], which incorporates a pretrained DINOv3 [28] vision backbone for robust visual feature extraction, to predict camera poses and dense 3D point maps. Since VGGT produces a dense point map $\mathbf{P}_i \in \mathbb{R}^{H\times W\times 3}$ with per-pixel confidence $\mathbf{w}_i$ for each view, we construct the initial point cloud by aggregating high-confidence predictions:

$$\mathcal{P} = \bigcup_{i=1}^{S} \{\mathbf{P}_i(x,y) \mid \mathbf{w}_i(x,y) > \tau_{\text{conf}}\}, \tag{8}$$

where τ_{conf} is a confidence threshold filtering unreliable predictions. The aggregated point cloud $\mathcal{P}$ and the estimated poses $\{\mathbf{T}_i\}_{i=1}^{S}$ are then employed to initialize 3DGS model for the subsequent training stage.

D. Deblurring 3DGS with Learnable Pose Refinement

Although the above two stages provide a point cloud with camera poses for 3DGS initialization, the inevitable pose errors may be accumulated during optimization and limits the ultimate performance. To remedy this, we jointly optimize 3D Gaussians and camera trajectories on the $SE(3)$ manifold, as shown in Fig.3.

1) Lie-Algebraic Pose Parameterization: For each blurry frame i, we introduce a learnable pose state $\mathcal{S}_i = (\mathbf{T}_i, \boldsymbol{\varpi}_i)$, where $\mathbf{T}_i \in SE(3)$ is the central pose and $\boldsymbol{\varpi}_i \in \mathfrak{se}(3)$ is the instantaneous motion during exposure. Pose updates are conducted via local perturbations [29] $\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{\nu})^\top \in \mathfrak{se}(3)$ through left multiplication $\mathbf{T}_i \leftarrow \exp(\boldsymbol{\xi}) \cdot \mathbf{T}_i$, avoiding singularities of global parameterizations.

2) Subframe Pose Construction: Given the pose state, N subframe poses $\{\mathbf{T}_{i,n}\}_{n=1}^{N}$ are constructed by interpolating the motion along the exposure interval:

$$\mathbf{T}_{i,n} = \exp(\delta_n \cdot \boldsymbol{\varpi}_i) \cdot \mathbf{T}_i, \quad \delta_n \in \left[-\frac{\Delta t}{2}, \frac{\Delta t}{2}\right] \tag{9}$$

where δ_n is the time offset for the n-th subframe. This formulation generates physically consistent subframe poses centered at $\mathbf{T}_i$.

3) Differentiable Rendering with Pose Gradients: To calculate the gradients of pose parameters within the CUDA rasterization pipeline, we reparameterize the rendering process. Mathematically, rendering a Gaussian at $\boldsymbol{\mu}$ with refined pose $\mathbf{T}_{\text{refined}}$ is equivalent to rendering a displaced Gaussian at $\boldsymbol{\mu}'$ with the original pose:

$$\boldsymbol{\mu}' = \mathbf{R}_{\text{orig}}^\top (\mathbf{R}_{\text{refined}}\boldsymbol{\mu} + \mathbf{t}_{\text{refined}} - \mathbf{t}_{\text{orig}}). \tag{10}$$

Since $\boldsymbol{\mu}'$ is differentiable w.r.t. the pose perturbation $\boldsymbol{\xi}$, this enables end-to-end pose optimization of camera poses.

4) Training Objective: Given Gaussians $\mathcal{G}$ and subframe poses $\{\mathbf{T}_n\}_{n=1}^{N}$, blurry images are synthesized by averaging subframe renderings:

$$\hat{\mathbf{B}} = \frac{1}{N}\sum_{n=1}^{N} \mathcal{R}(\mathcal{G}, \mathbf{T}_n), \tag{11}$$

where $\mathcal{R}(\cdot)$ is the differentiable 3DGS renderer. The photometric loss is defined as:

$$\mathcal{L}_{\text{photo}} = (1-\lambda)\|\hat{\mathbf{B}} - \mathbf{B}\|_1 + \lambda\mathcal{L}_{\text{D-SSIM}}(\hat{\mathbf{B}}, \mathbf{B}) \tag{12}$$

where the weight $\lambda = 0.2$ balances the L1 loss and structural dissimilarity loss.

5) Alternating Optimization: Jointly optimizing geometry $\mathcal{G}$ and camera poses $\mathcal{S} = \{(\mathbf{T}_i, \boldsymbol{\varpi}_i)\}$ is ill-posed due to the inherent geometry-motion ambiguity. We therefore train our framework using an alternating optimization strategy

(i) Gaussian Update. With poses $\mathcal{S}$ frozen, Gaussian attributes $\boldsymbol{\theta}_{\mathcal{G}}$ are updated:

$$\boldsymbol{\theta}_{\mathcal{G}}^{(k+1)} = \arg\min_{\boldsymbol{\theta}_{\mathcal{G}}} \mathcal{L}_{\text{photo}}(\boldsymbol{\theta}_{\mathcal{G}}, \mathcal{S}^{(k)}) \tag{13}$$

(ii) Pose Update. With Gaussians $\boldsymbol{\theta}_{\mathcal{G}}$ frozen, pose states $\mathcal{S}$ are updated:

$$\mathcal{S}^{(k+1)} = \arg\min_{\mathcal{S}} \mathcal{L}_{\text{photo}}(\boldsymbol{\theta}_{\mathcal{G}}^{(k+1)}, \mathcal{S}) \tag{14}$$

After each pose update, manifold retraction $\mathbf{T}_i \leftarrow \exp(\boldsymbol{\xi}_i^*)\cdot\mathbf{T}_i$ is applied and the perturbation is reset ($\boldsymbol{\xi}_i \leftarrow \mathbf{0}$) to maintain numerical stability.

V. Experiments

A. Datasets and Implementation Details

To comprehensively evaluate our method, we conduct experiments on the following datasets, covering synthetic, real-world, and self-collected scenarios.

Synthetic Dataset. We adopt 8 standard scenes from the NeRF Synthetic dataset [30] (Lego, Drums, Materials, Ficus, Chair, Hotdog, Mic, and Ship). To simulate realistic motion blur, we follow the protocol in Deblur-NeRF [17]. For each view, we generate a smooth camera trajectory and render 20 sub-frames, which are then averaged to produce the final blurry image. Corresponding event streams are simulated between the sub-frames using the event camera simulator ESIM [31].

Real Scene Datasets. We evaluate on two real scene datasets. First, we adopt the *Truck* and *Train* scenes from the Real Scene Benchmark [1], where motion blur and event streams are synthesized following the same sub-frame averaging and ESIM simulation protocol as the synthetic dataset. Second, we adopt sequences from the DAVIS240C event camera dataset [32], which provides synchronized grayscale frames and event streams captured by a DAVIS240C sensor at a resolution of 240×180 pixels. This dataset features various camera motion patterns including translational, rotational, and combined 6-DOF motions in indoor environments, with ground truth poses provided by a motion capture system.

Self-Collected Dataset. To validate our method on real-world data, we build a UAV-mounted multi-sensor platform equipped with a DAVIS346 event camera and a standard RGB

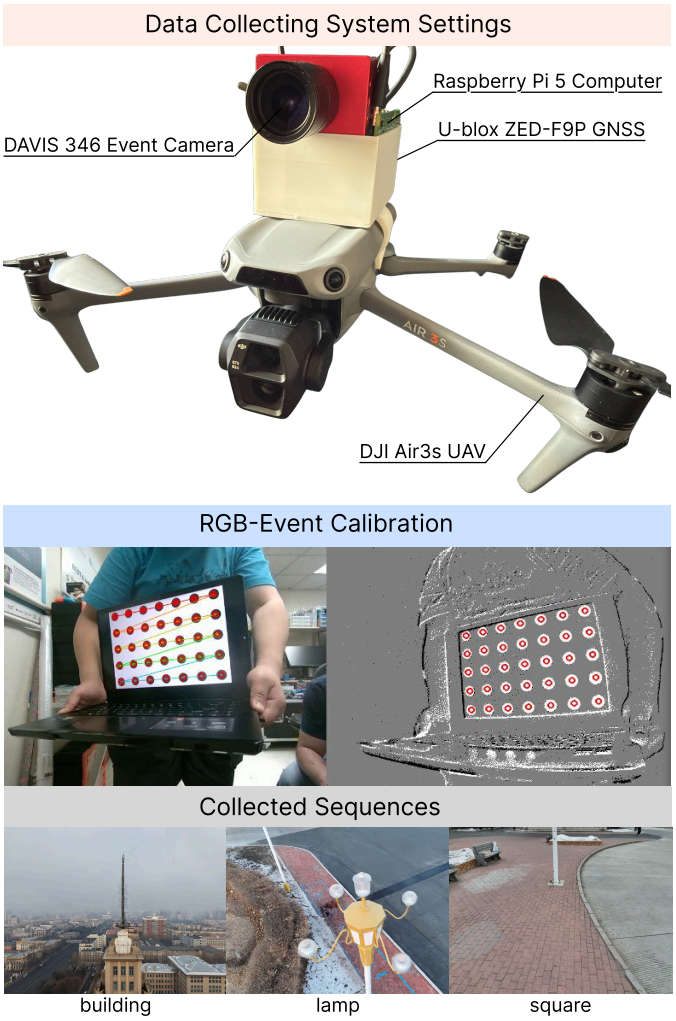

Fig. 4. Our UAV-mounted data acquisition platform. A DAVIS346 event camera and a standard RGB camera are rigidly mounted on the UAV frame and calibrated using a blinking chessboard pattern. Three sequences are collected: *Building* (large-scale environment), *Square* (featureless and repetitive scene), and *Lamp* (fine outdoor object).

camera, as shown in Fig. 4. The two sensors are rigidly mounted and calibrated using a blinking chessboard pattern displayed on a screen. We employ the hardware synchronization method from TEVIO [5] to strictly align the RGB frames and event streams. We collect three sequences targeting diverse and challenging scenarios: *Building*, which captures a large-scale outdoor environment with rich geometric structures; *Square*, which features a featureless and repetitive scene with duplicated architectural elements that challenges conventional SfM methods; and *Lamp*, which focuses on precise modeling of an outdoor lamp object with fine geometric details.

Implementation Details. During training, we use the Adam optimizer with standard parameters ($\beta_1 = 0.9, \beta_2 = 0.999$) to train our framework for 30,000 iterations. Within the learnable pose refinement module, we set the learning rates for rotation and translation to 1×10^{-3} and 1×10^{-4} with an exponential decay scheduler. All experiments are conducted on a single NVIDIA RTX 4090 GPU.

B. Comparison with State-of-the-Art Approaches

We compare our method with three categories of baselines: 1) B-3DGS [1], the vanilla 3D Gaussian Splatting directly trained on blurry images. 2) BAD-NeRF [33], an RGB-only deblurring method that physically models the blur formation process and jointly optimizes camera trajectories via bundle adjustment without utilizing event data. 3) Event-aided methods, including ED-NeRF [7], Event-3DGS [14], and E^{2}NeRF [6], which incorporate event streams to assist the neural rendering process under motion blur.

Evaluation Metrics. For datasets with ground truth sharp images (NeRF Synthetic, Real Scene Benchmark, and DAVIS240C), we employ three full-reference metrics PSNR, SSIM [34], and LPIPS [35] to evaluate novel view rendering quality. For the Self-Collected dataset where sharp ground truth images are unavailable, we adopt five no-reference image quality assessment metrics, including BRISQUE [36], NIQE [37], PIQE, RankIQA, and MetaIQA.

Quantitative Results. Table I reports the quantitative comparisons across all four datasets. It can be observed that our method achieves the best performance across all metrics on all datasets.

On the NeRF Synthetic dataset, B-3DGS produces inferior results (22.55 dB) since it is directly trained on blurred images without any deblurring mechanism. BAD-NeRF improves upon B-3DGS by 0.63 dB through its bundle-adjusted blur modeling; however, it still suffers from inaccurate pose initialization under severe blur due to its dependence on RGB-based SfM. The event-aided methods (ED-NeRF, Event-3DGS, E^{2}NeRF) produce substantial improvements by leveraging blur-free event observations, with E^{2}NeRF reaching 25.12 dB. Nevertheless, these methods either adopt indirect initialization strategies or keep camera poses fixed during optimization. In comparison, our EEDGS benefits from the direct event-based initialization and the joint pose-geometry optimization to outperfrom E^{2}NeRF by 1.22 dB.

On the Real Scene Benchmark, a similar trend can be observed. Specifically, our method achieves 24.56 dB, surpassing Event-3DGS by 1.11 dB. Notably, the performance gap between B-3DGS (20.12 dB) and Event-3DGS (23.45 dB) reaches 3.33 dB, indicating that leveraging event data for initialization becomes increasingly critical as the scene complexity grows.

On the DAVIS240C dataset with real event camera data, our method achieves 31.53 dB, demonstrating an improvement of 1.66 dB over E^{2}NeRF (29.87 dB). The higher overall performance compared to the synthetic benchmarks can be attributed to the availability of genuine event data rather than simulated events, which provides more accurate and temporally dense motion cues for our event-based initialization.

For the Self-Collected dataset, we adopt NR-IQA metrics since ground truth images are unavailable. Compared to the second-best method, our method produces a 14.61% gain in BRISQUE, a 12.34% gain in NIQE, a 16.50% gain in PIQE, a 10.59% gain in RankIQA, and a 17.14% gain in MetaIQA, demonstrating superior perceptual quality in challenging real-world UAV scenarios.

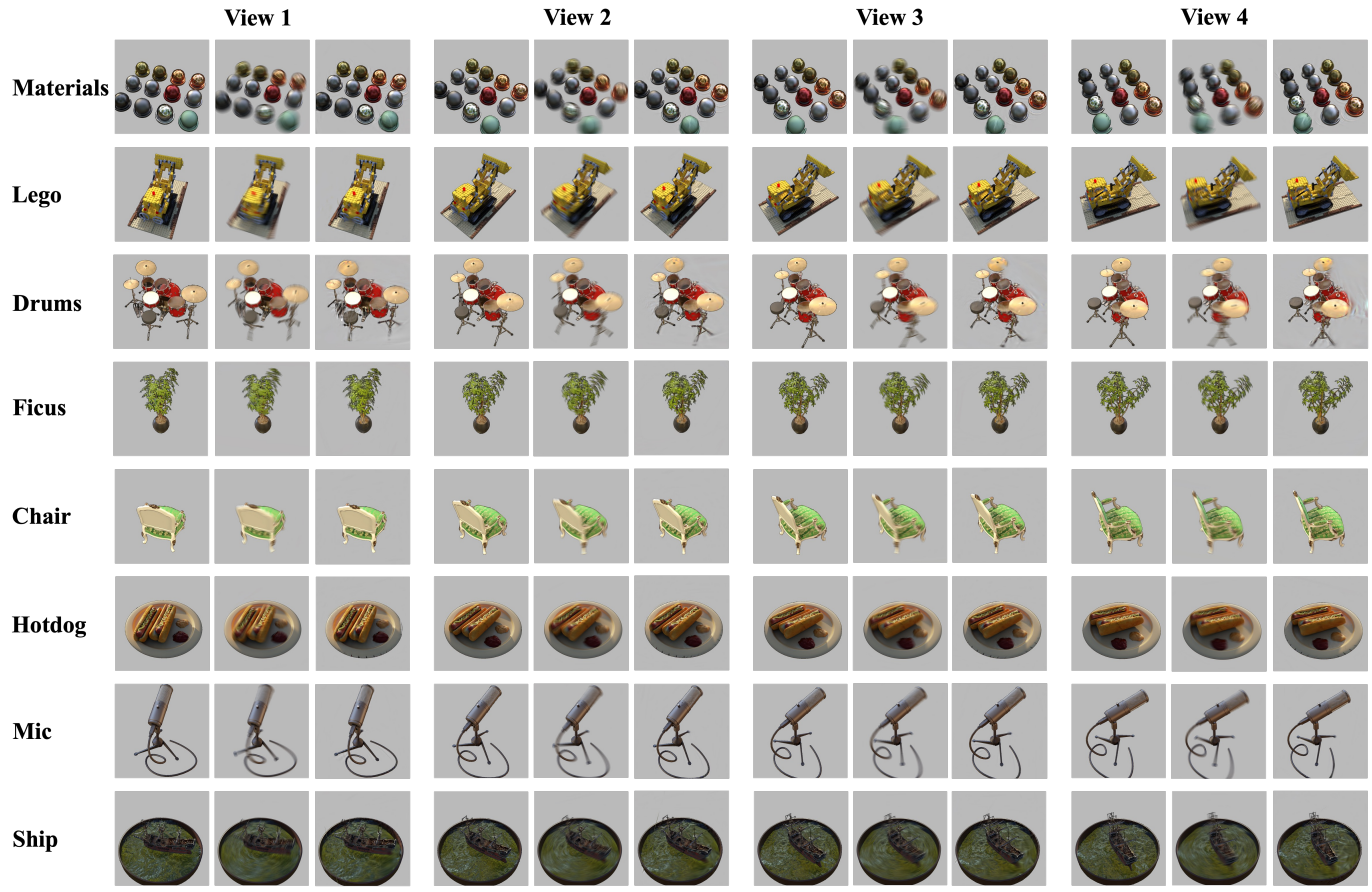

Fig. 5. Qualitative results of our EEDGS on the NeRF Synthetic dataset. Each row corresponds to a different scene. For each scene, we show 4 novel test views, where each group of three columns displays: sharp ground truth (left), blurry input (middle), and our deblurred result (right).

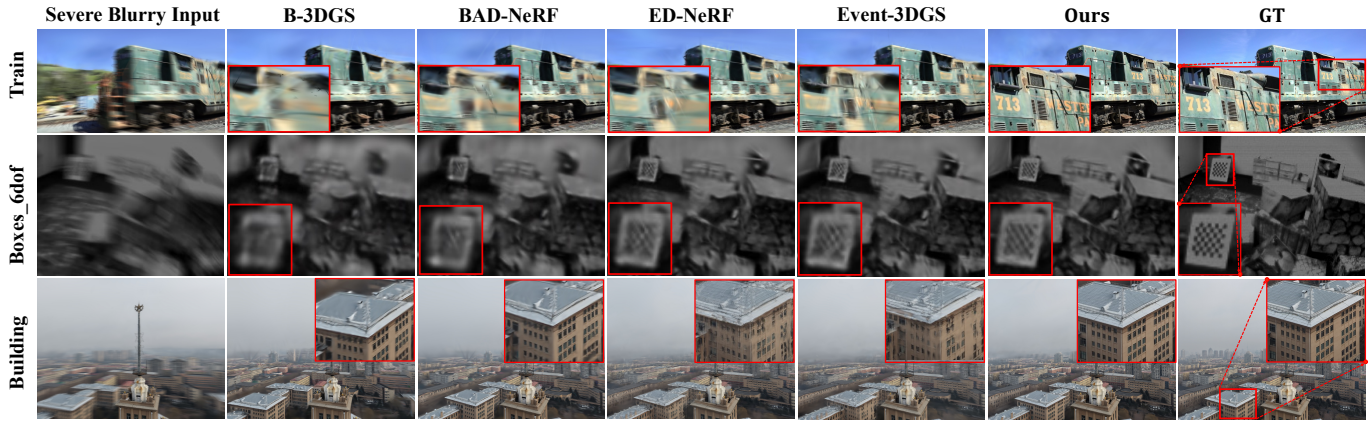

Fig. 6. Visual comparison with baseline methods on three representative real-world scenes. From top to bottom: the Real Scene Benchmark, the DAVIS240C dataset [32], and our Self-Collected dataset.

Qualitative Results. Fig. 5 presents qualitative results of our EEDGS on all 8 scenes from the NeRF Synthetic dataset. Fig. 6 further provides visual comparisons with all baseline methods on three representative real-world scenes. B-3DGS produces heavily blurred novel views, and BAD-NeRF recovers partial structures but still exhibits noticeable artifacts in texture-rich regions. The event-aided methods produce overall satisfactory results but struggle to recover fine-grained details due to fixed camera poses or over-smoothing. In comparison, our method produces high-fidelity renderings with sharp edges and intricate textures, attributed to the synergy between event-based initialization and the learnable pose refinement module.

C. Camera Trajectory Accuracy

A key contribution of our method is the ability to recover accurate camera trajectories from blurry inputs. Fig. 7 visual-

TABLE I
QUANTITATIVE COMPARISON ACROSS DIFFERENT DATASETS. BEST RESULTS ARE IN **BOLDFACE** WHILE SECOND BEST RESULTS ARE UNDERLINED. ↑: HIGHER IS BETTER, ↓: LOWER IS BETTER.

Type	Dataset	Metric	B-3DGS	BAD-NeRF	ED-NeRF	Event-3DGS	E^{2}NeRF	Ours
Synthetic	NeRF Synthetic	PSNR ↑	22.55	23.18	24.23	24.67	25.12	**26.34**
		SSIM ↑	0.852	0.861	0.878	0.891	0.872	**0.905**
		LPIPS ↓	0.192	0.185	0.162	0.148	0.175	**0.138**
Real-world	Real Scene Benchmark	PSNR ↑	20.12	21.03	22.05	23.45	22.34	**24.56**
		SSIM ↑	0.812	0.825	0.838	0.841	0.853	**0.873**
		LPIPS ↓	0.235	0.221	0.198	0.182	0.201	**0.162**
	DAVIS240C	PSNR ↑	21.34	23.56	28.92	29.45	29.87	**31.53**
		SSIM ↑	0.712	0.768	0.870	0.880	0.875	**0.910**
		LPIPS ↓	0.342	0.285	0.175	0.168	0.171	**0.153**
	Self-Collected	BRISQUE ↓	68.45	63.21	55.78	48.67	54.23	**41.56**
		NIQE ↓	14.56	13.12	11.89	11.23	10.78	**9.45**
		PIQE ↓	72.34	66.45	56.89	51.23	58.67	**42.78**
		RankIQA ↑	5.23	5.45	5.89	6.23	5.78	**6.89**
		MetaIQA ↑	0.178	0.195	0.225	0.218	0.245	**0.287**

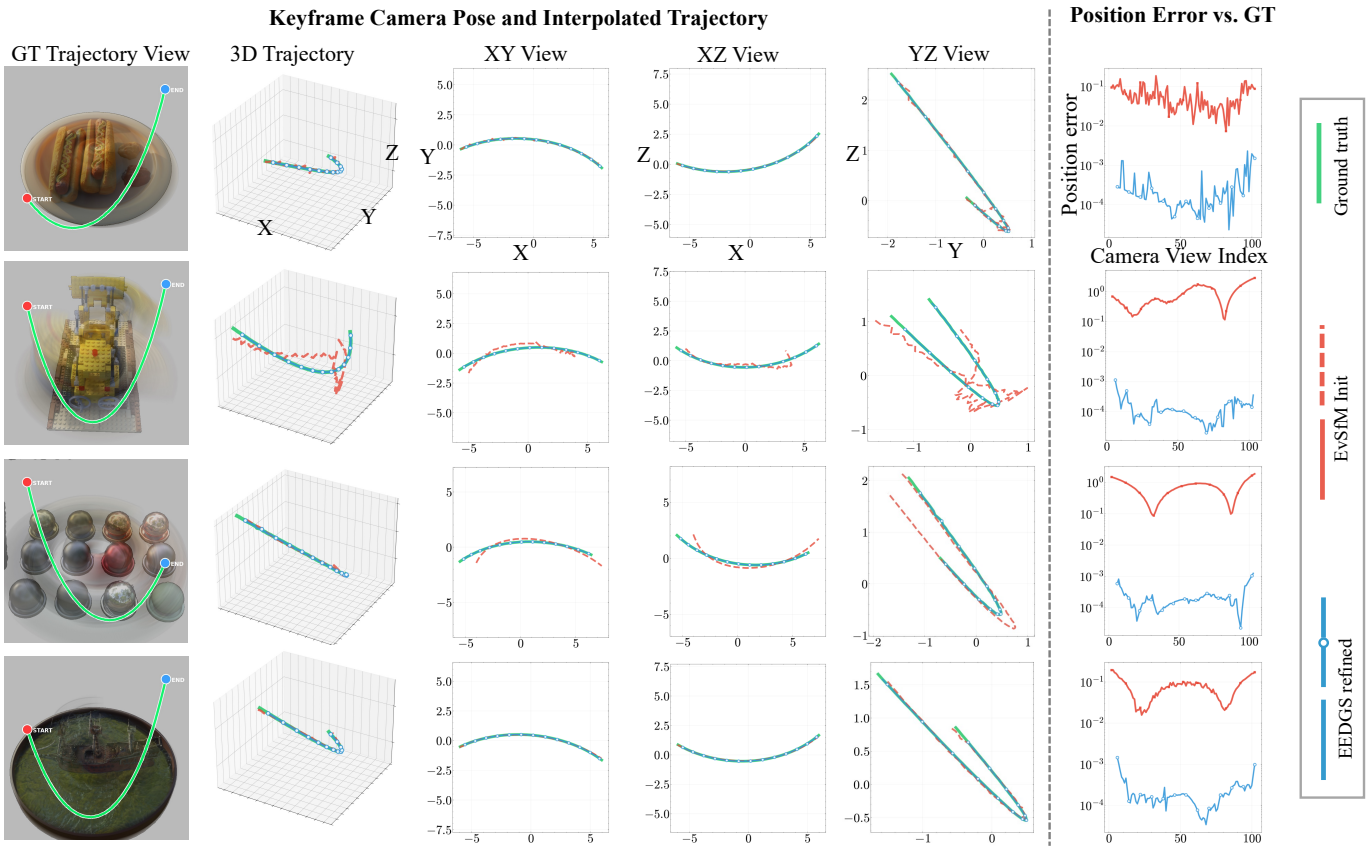

Fig. 7. Camera trajectory comparison on four representative synthetic scenes. For each scene, we show the 3D trajectory and three orthogonal projections comparing ground truth, event-based initialization (EvSfM Init), and our refined poses, along with per-frame position error magnitude.

izes the estimated trajectories on four representative synthetic scenes. It can be observed that the event-based initialization provides trajectories that roughly align with the ground truth. Furthermore, the learnable pose refinement module consistently reduces the per-frame position error across all scenes, producing trajectories that align closer to the ground truth. This confirms that jointly optimizing camera poses with scene geometry on the $SE(3)$ manifold effectively corrects residual initialization errors and improves geometric consistency.

D. Ablation Studies

To validate the contribution of each core component in our framework, we conduct ablation studies with quantitative rendering quality evaluation on the NeRF Synthetic dataset

TABLE II
ABLATION STUDY OF KEY COMPONENTS ON SELF-COLLECTED DATASET.

Method Setting	PSNR ↑	SSIM ↑	LPIPS ↓
w/o Event Init (RGB SfM)	20.15	0.782	0.241
w/o Pose Refinement	24.67	0.891	0.148
w/o Event Img. Norm.	25.12	0.895	0.142
Full Model (EEDGS)	**26.34**	**0.905**	**0.138**

(Table II) and qualitative camera trajectory analysis on our self-collected dataset (Fig. 8).

Event-Based Initialization. The event-based initialization module infers camera pose and point cloud directly from event streams, bypassing blur-corrupted RGB images to provide a reliable starting point for 3DGS training. To evaluate its contribution, we replace it with the standard RGB-based COLMAP pipeline, denoted as "w/o Event Init." As shown in Table II, the PSNR drops drastically from 26.34 dB to 20.15 dB (-6.19 dB), with SSIM degrading from 0.905 to 0.782 and LPIPS increasing from 0.138 to 0.241. These results indicate that the event-based initialization is the most critical component in EEDGS, as it circumvents the catastrophic failure of RGB-based SfM under severe motion blur and provides the foundation for all subsequent optimization stages.

Learnable Pose Refinement. The learnable pose refinement module jointly optimizes camera trajectories and scene geometry on the $SE(3)$ manifold, compensating for pose errors produced during the event-based initialization stage. To analyze its impact, we freeze all camera poses after initialization and optimize only the Gaussian attributes, denoted as "w/o Pose Refinement". The PSNR drops from 26.34 dB to 24.67 dB (-1.67 dB), and LPIPS increases from 0.138 to 0.148. This demonstrates that even with robust event-based initialization, the pose errors could be accumulated during the optimization of 3DGS to limit the ultimate reconstruction quality. In contrast, our learnable pose refinement module can effectively correct these errors through alternating optimization to produce notable gains. Fig. 8 further visualizes this effect on our self-collected dataset. without pose refinement, the estimated trajectory exhibits noticeable drift, whereas the refined trajectory closely aligns with the ground truth.

Event Image Normalization. The event image normalization applies a learnable affine transformation to align event image statistics with natural images, facilitating more accurate predictions by VGGT. Removing this normalization and directly feeding standardized event images to VGGT ("w/o Event Img. Norm.") decreases PSNR from 26.34 dB to 25.12 dB (-1.22 dB) and SSIM from 0.905 to 0.895, confirming that bridging the domain gap improves initial pose accuracy and propagates as better reconstruction quality.

VI. CONCLUSION

In this paper, we propose EEDGS, a unified framework that exploits event-based geometry cues for robust 3D Gaussian Splatting under severe motion blur. Our method first extracts geometric priors, including camera poses and scene point clouds, directly from event streams, bypassing blur-corrupted RGB-based SfM entirely. Building upon these event-derived geometric priors, a differentiable trajectory optimization module jointly refines camera poses with scene geometry during 3DGS training. Extensive experiments on synthetic and real-world datasets demonstrate that EEDGS achieves state-of-the-art performance in both rendering quality and pose estimation accuracy. Future work will explore extending this framework to dynamic scenes and real-time online reconstruction.

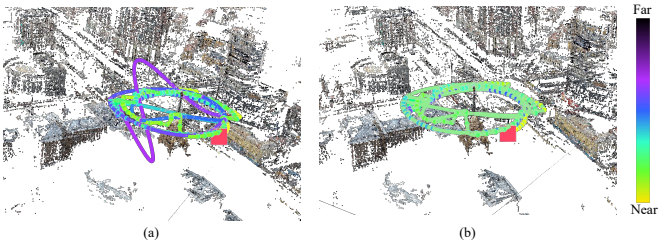

Fig. 8. Estimated camera trajectories on self-collected data. (a) Resutls produced by our method without the learnable pose refinement module. (b) Results produced by our full model. Trajectory color indicates pose error.

REFERENCES

[1] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for real-time radiance field rendering," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 1–14, 2023.
[2] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4104–4113, 2016.
[3] B. Lee, H. Lee, X. Sun, U. Ali, and E. Park, "Deblurring 3D Gaussian Splatting," in *Proc. Eur. Conf. Comput. Vis.*, pp. 127–143, 2024.
[4] L. Zhao, P. Wang, and P. Liu, "BAD-Gaussians: Bundle adjusted deblur Gaussian Splatting," in *Proc. Eur. Conf. Comput. Vis.*, pp. 233–250, 2024.
[5] G. Gong, F. Hu, F. Wang, M. Muddassir, P. Zhou, L. Li, Q. Wang, Z. He, and D. Navarro-Alarcon, "TEVIO: Thermal-aided event-based visual inertial odometry for robust state estimation in challenging environments," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1124–1138, 2025.
[6] Y. Qi, L. Zhu, Y. Zhang, and J. Li, "E2NeRF: Event enhanced neural radiance fields from blurry images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 13254–13264, 2023.
[7] M. Cannici and D. Scaramuzza, "Mitigating motion blur in neural radiance fields with events and frames," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 9286–9296, 2024.
[8] 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.
[9] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and R. Scaramuzza, "Event-based vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, 2020.
[10] I. Hwang, J. Kim, and Y. M. Kim, "Ev-NeRF: Event based neural radiance field," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, pp. 837–847, 2023.
[11] S. Klenk, L. Koestler, D. Scaramuzza, and D. Cremers, "E-NeRF: Neural radiance fields from a moving event camera," *IEEE Robot. Autom. Lett.*, vol. 8, no. 3, pp. 1587–1594, 2023.
[12] V. Rudnev, M. Elgharib, C. Theobalt, and V. Golyanik, "EventNeRF: Neural radiance fields from a single colour event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 4992–5002, 2023.
[13] W. F. Low and G. H. Lee, "Robust e-NeRF: NeRF from sparse & noisy events under non-uniform motion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 18335–18346, 2023.
[14] H. Han, J. Li, H. Wei, and X. Ji, "Event-3DGS: Event-based 3D reconstruction using 3D Gaussian Splatting," in *Adv. Neural Inf. Process. Syst.*, vol. 37, pp. 128139–128159, 2024.

[15] W. Yu, C. Feng, J. Li, J. Tang, J. Yang, Z. Tang, M. Cao, X. Jia, Y. Yang, L. Yuan, *et al.*, "EvaGaussians: Event stream assisted Gaussian Splatting from blurry images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 24780–24790, 2025.
[16] T. Yura, A. Mirzaei, and I. Gilitschenski, "EventSplat: 3D Gaussian Splatting from moving event cameras for real-time rendering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 26876–26886, 2025.
[17] L. Ma, X. Li, J. Liao, Q. Zhang, X. Wang, J. Wang, and P. V. Sander, "Deblur-NeRF: Neural radiance fields from blurry images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 12861–12870, 2022.
[18] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, "BARF: Bundle-adjusting neural radiance fields," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 5741–5751, 2021.
[19] W. Bian, Z. Wang, K. Li, J.-W. Bian, and V. A. Prisacariu, "NoPe-NeRF: Optimising neural radiance field with no pose prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 4160–4169, 2023.
[20] Y. Fu, S. Liu, A. Kulkarni, J. Kautz, A. A. Efros, and X. Wang, "COLMAP-free 3D Gaussian Splatting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 20796–20805, 2024.
[21] A. Meuleman, Y.-L. Liu, C. Gao, J.-B. Huang, C. Kim, M. H. Kim, and J. Kopf, "Progressively optimized local radiance fields for robust view synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 16539–16548, 2023.
[22] H. Matsuki, R. Murai, P. H. J. Kelly, and A. J. Davison, "Gaussian Splatting SLAM," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 18039–18048, 2024.
[23] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, and J. Revaud, "DUSt3R: Geometric 3D vision made easy," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 20697–20709, 2024.
[24] V. Leroy, Y. Cabon, and J. Revaud, "Grounding image matching in 3D with MASt3R," in *Proc. Eur. Conf. Comput. Vis.*, pp. 71–91, 2024.
[25] L. Pan, C. Scheerlinck, X. Yu, R. Hartley, M. Liu, and Y. Dai, "Bringing a blurry frame alive at high frame-rate with an event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 6820–6829, 2019.
[26] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 3867–3876, 2018.
[27] J. Wang, M. Chen, N. Karaev, A. Vedaldi, C. Rupprecht, and D. Novotny, "VGGT: Visual Geometry Grounded Transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 5294–5306, 2025.
[28] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, *et al.*, "DINOv3," *arXiv preprint arXiv:2508.10104*, 2025.
[29] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Image Vis. Comput.*, vol. 30, no. 2, pp. 65–77, 2012.
[30] 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," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, 2021.
[31] H. Rebecq, D. Gehrig, and D. Scaramuzza, "ESIM: An open event camera simulator," in *Conf. Robot. Learn.*, pp. 969–982, 2018.
[32] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Res.*, vol. 36, no. 2, pp. 142–149, 2017.
[33] P. Wang, L. Zhao, R. Ma, and P. Liu, "BAD-NeRF: Bundle adjusted deblur neural radiance fields," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 4170–4179, 2023.
[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
[35] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 586–595, 2018.
[36] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
[37] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2012.

Gu Gong received the M.Sc. degree in control science and technology from the Harbin Institute of Technology (HIT), Harbin, China, in 2019. He is pursuing a Ph.D. in mechanical engineering at The Hong Kong Polytechnic University and a Ph.D. in control science and engineering at the HIT. His research interests focus on event-based vision, multi-sensor fusion, and SLAM.

Fangyuan Wang received the M.Sc. degree in software engineering from Zhejiang Sci-Tech University, China, in 2022. He is currently pursuing the Ph.D. in mechanical engineering at The Hong Kong Polytechnic University, Hong Kong. His research interests focus on reinforcement learning, multi-agent systems, and robotic manipulation.

Zhen He received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2000. From 1997 to 1998, she was a visiting Ph.D. student at the Mita Laboratory of Tokyo Institute of Technology, Japan. She has been with Harbin Institute of Technology since 2000 and is currently a Professor with the Department of Control Science and Engineering. Her research interests include robust control, optimal control, H-infinity control, and generalized systems.

Qiang Wang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1998, 2000, and 2004, respectively. Since 2008, he has been a professor at the Department of Control Science and Engineering, HIT. His research interests include hyperspectral image denoising, signal/image processing, multi-sensor data fusion, wireless sensor networks, and intelligent detection technology.

David Navarro-Alarcon (Senior Member, IEEE) received the Ph.D. degree in mechanical and automation engineering from The Chinese University of Hong Kong (CUHK), in 2014. From 2014 to 2017, he worked as a Postdoctoral Fellow and then as a Research Assistant Professor at the T Stone Robotics Institute of CUHK. Since 2017, he has been with The Hong Kong Polytechnic University (PolyU), where he is currently an Associate Professor with the Department of Mechanical Engineering, and the Principal Investigator of the Robotics and Machine Intelligence Laboratory. His current research interests include perceptual robotics and control systems. Dr. Navarro-Alarcon currently serves as an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS.