EvHDR-GS: Event-guided HDR Video Reconstruction with 3D Gaussian Splatting

Zehao Chen[1,2*], Zhan Lu[3*], De Ma[1,2], Huajin Tang[1,2], Xudong Jiang[3], Qian Zheng[1,2†], Gang Pan[1,2†]

[1] The State Key Lab of Brain-Machine Intelligence, Zhejiang University
[2] College of Computer Science and Technology, Zhejiang University
[3] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
{zehao, made, htang, qianzheng, gpan}@zju.edu.cn , {zhan007, exdjiang}@ntu.edu.sg

Abstract

High Dynamic Range (HDR) video reconstruction seeks to accurately restore the extensive dynamic range present in real-world scenes and is widely employed in downstream applications. Existing methods typically operate on one or a small number of consecutive frames, which often leads to inconsistent brightness across the video due to their limited perspective on the video sequence. Moreover, supervised learning-based approaches are susceptible to data bias, resulting in reduced effectiveness when confronted with test inputs exhibiting a domain gap relative to the training data. To address these limitations, we present an event-guided HDR video reconstruction method through building 3D Gaussian Splatting (3DGS), to ensure consistent brightness imposed by 3D consistency. We introduce HDR 3D Gaussians capable of simultaneously representing HDR and low-dynamic-range (LDR) colors. Furthermore, we incorporate a learnable HDR-to-LDR transformation optimized by input event streams and LDR frames to eliminate the data bias. Experimental results on both synthetic and real-world datasets demonstrate that the proposed method achieves state-of-the-art performance.

Introduction

The human visual system perceives a broader dynamic range to perceptual the environment. However, limited by the low dynamic range (LDR) of conventional cameras, the recorded LDR video might lose high dynamic range (HDR) information, *e.g.*, vital highlight and shadow details. As HDR information is useful for improving realism, enhances immersion, and supports a range of applications in photography, cinematography, video games, and virtual reality, HDR video reconstruction tasks become more significant.

Existing HDR reconstruction methods focus on recovering frame-by-frame HDR information for LDR video, which could be categorized into two groups by different input data requirements, *i.e.*, multi-exposed and single-exposed inputs. However, each of them has its challenges: 1) multi-exposure methods suffer from complex capturing settings and ghosting artifacts (Chen et al. 2021a; Yang et al. 2023), 2) single-exposure methods are inherently ill-posed due to the lack of physical cues for reconstructing HDR information (Yang et al. 2023). Recent advances leverage the event camera to perform HDR reconstruction (Han et al. 2020; Yang et al.

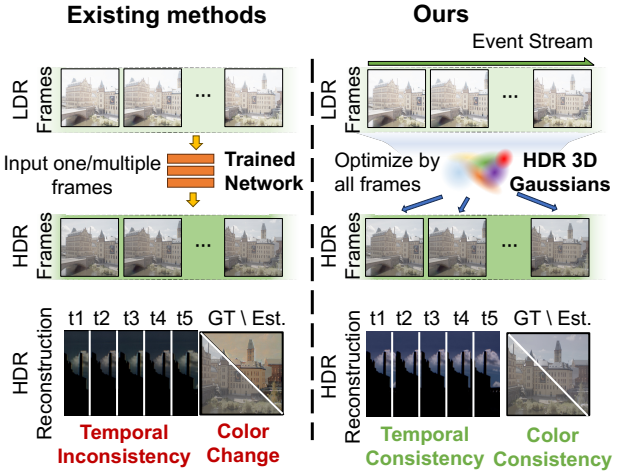

Figure 1: Comparison of existing HDR video reconstruction methods (left) and our method (right). Existing approaches frame-by-frame reconstruct HDR information for one or multiple input LDR frames through pre-trained neural networks. In contrast, our method employs scene representation 3D Gaussian Splatting (Kerbl et al. 2023) to represent entire LDR frames in the video and conduct HDR reconstruction based on the modeling of the relationship between RGB frames and events. Compared to the results of HDR reconstruction, 1) the frame-based method (Chung and Cho 2023) stitches 6 consecutive LDR frames for reducing the exposure time, while it suffers from brightness inconsistencies between HDR frames (*e.g.*, t1 to t5). In contrast, our method produces more temporal consistent HDR results; 2) the event-guided method (Yang et al. 2023) leverages a neural network trained in a supervised manner to estimate HDR frames, while it introduces data bias and results in a yellowish color tint. In contrast, our method adopts a self-supervised approach, learning directly from the distributions in the current video, thereby avoiding issues related to data bias.

*These authors contributed equally.
†Corresponding author.

2023; Cui et al. 2024b), since it could simply the input requirements and naturally capture HDR information. Nevertheless, both frame-based methods and event-guided methods could not perfectly address the challenge of HDR video reconstruction as they suffer from **brightness inconsistencies** between estimated adjacent HDR frames (see example in Figure 1(b)). Although recent efforts employ temporal constraints to enhance consistency and reduce flickering (Chung and Cho 2023; Yang et al. 2023), they could not fully eliminate this issue since their frame-by-frame mechanism intrinsically processes frames independently. Besides, all event-guided methods and most frame-based methods are conducted in a supervised manner, *i.e.*, training the networks with pre-collected data (pairs of HDR frames and LDR frames, or with aligned events). These supervised methods are affected by **data bias** from collected training data through training procedure, which might degrade their performance dramatically when the test cases have a large domain gap with the training ones. For example, as illustrated in Figure 1(c), the event-guided method often produces a yellowish tint in the reconstructed HDR videos due to training data bias. These two primary problems limit the effectiveness of existing methods to produce faithful and consistent HDR video from LDR frames.

Motivated by existing scene reconstruction methods (*e.g.*, 3D Gaussian Splatting (3DGS) (Kerbl et al. 2023)) and their optimization processes, we transfer the HDR video reconstruction task to an HDR scene reconstruction task and optimized only with input LDR frames and captured events, to ensure temporal brightness consistency and alleviate the training data bias, respectively. Specifically, we treat the entire LDR video frames as multiple views to build an HDR scene, thus the rendered HDR images corresponding to video frames are inherently temporal consistent due to the 3D consistency. Besides, we model an LDR-to-HDR transformation for the input LDR frames and estimated HDR frames, and learn it from the paired event information in a self-supervised manner to avoid data bias.

To this end, we propose a self-supervised **Ev**ent-guided **HDR** video reconstruction method through 3D**GS**, namely EvHDR-GS. After building an HDR scene, the reconstructed geometry imposes the temporal brightness consistency of each HDR frame estimation. We select 3DGS as the scene representation due to its fast optimization and add HDR output branch into its original LDR output. The proposed method only requires single-exposed LDR frames as inputs with the help of paired events. The two branches of 3DGS outputs are constrained with a proposed LDR-to-HDR transformation and learned from the paired frames and events. Extensive experiments on both synthetic and real-world scenarios demonstrate that the proposed EvHDR-GS achieves state-of-the-art performance compared to previous works. The main contributions can be summarized as

1. We utilize a scene representation to model LDR video frames and transfer the video HDR reconstruction task to HDR scene reconstruction work to ensure temporal consistency.
2. We propose a self-supervised method for HDR video reconstruction using events stream and single-exposure LDR frames, to eliminate the data bias by learning the LDR-to-HDR transformation from their relationship.
3. We demonstrate that our method achieves state-of-the-art performance by producing accurate and consistent HDR results on both synthetic and real-world datasets, validating its effectiveness on video HDR reconstruction.

Related Work

Frames-based HDR video reconstruction. Frames-based HDR video reconstruction methods can be divided into two categories: multi-exposure methods and single-exposure methods. Multi-exposure methods merge a stack of LDR frames with multi-exposure into HDR frames. The multi-exposure LDR frames can come from different sensors (Nayar and Mitsunaga 2000; Tocci et al. 2011) or different periods of the same sensor (Kang et al. 2003; Kalantari and Ramamoorthi 2019; Chen et al. 2021a; Chung and Cho 2023). Due to motion during video capture, this type of method suffers from ghosting artifacts. Recently, a series of approaches (Yan et al. 2019; Liu et al. 2022; Song et al. 2022; Chung and Cho 2023; Shu et al. 2024; Xu et al. 2024; Cui et al. 2024a) have been developed that align multi-exposure frames by adding various modules. Single-exposure methods (Endo, Kanamori, and Mitani 2017; Rempel et al. 2007; Eilertsen et al. 2017) can also be used to address the issue of ghosting artifacts. However, these methods lack physical guidance in over/under-exposed regions, making it an ill-posed problem.

Both frame-based methods suffer from brightness inconsistencies between frames. This issue arises because these methods perform HDR reconstruction on one or a few adjacent frames at a time. Although temporal loss is added to enhance temporal consistency and reduce flickering (Chung and Cho 2023), these methods process the video in segments and cannot fully resolve brightness inconsistencies between consecutive frames. In contrast, our method uses 3DGS to represent the entire video, leveraging the multi-view consistency of 3DGS to eliminate brightness inconsistency issues.

Events-based HDR reconstruction. Event cameras, which are neuromorphic vision sensors, offer high temporal resolution and superior dynamic range compared to conventional cameras, making them valuable in various fields (Chen et al. 2021b, 2024; Lele et al. 2021; Li et al. 2018). Owing to these capabilities, researchers have increasingly employed them for HDR reconstruction tasks. A series of approaches (Kim et al. 2008; Bardow, Davison, and Leutenegger 2016; Wang et al. 2019; Rebecq et al. 2019a,b; Wang, Kim, and Yoon 2020; Mostafavi, Wang, and Yoon 2021; Zou et al. 2021) use only events for HDR reconstruction. However, since event cameras can only record intensity changes, this type of reconstruction is inherently ill-posed. Another category of work uses events as a guidance tool to convert LDR images into HDR images (Pan et al. 2019, 2020; Messikommer et al. 2022; Han et al. 2023; Samra, Mitra, and Shedligeri 2023; Li et al. 2024). (Han et al. 2020) applies events to HDR imaging by converting events to intensity maps, while (Cui

et al. 2024b) utilizes color events in single-exposure HDR imaging. (Yang et al. 2023) used events for HDR video reconstruction.

Compared to other video reconstruction methods, ours is self-supervised and considers the video as a whole, effectively addressing the issue of brightness inconsistency.

3D Gaussian Splatting (3DGS). 3DGS (Kerbl et al. 2023) employs millions of Gaussian point clouds for explicit scene representation. It features a highly parallelized rasterization workflow, which accelerates the view rendering process. This approach enables significantly faster inference speeds compared to NeRF-based methods (Mildenhall et al. 2021; Lu et al. 2024; Liao et al. 2024). Due to these advantages, 3DGS has been rapidly adopted and is now widely used in a variety of fields, including dynamic scene rendering (Yang et al. 2024; Lin et al. 2024; Wu et al. 2024; Huang et al. 2024; Sun et al. 2024), inverse rendering (Liang et al. 2024), and 3D reconstruction (Zhang et al. 2024).

Compared to other 3DGS-related methods, we are the first to apply 3DGS to HDR video reconstruction.

Preliminary

3D Gaussian Splatting. Previous works (Kerbl et al. 2023) propose to represent a 3D scene as a set of 3D Gaussian primitives $\{\mathcal{G}_k \mid k = 1, \cdots, K\}$ and employ volume splitting for image rendering. The geometry of each scaled 3D Gaussian $\mathcal{G}_k$ is parameterized as:

$$\mathcal{G}_k(\mathbf{x}) = \alpha_k e^{-\frac{1}{2}(\mathbf{x}-\mathbf{p}_k)^T \mathbf{\Sigma}_k^{-1}(\mathbf{x}-\mathbf{p}_k)}, \tag{1}$$

where an opacity $\alpha_k \in [0, 1]$, center $\mu \in \mathbb{R}^{3\times1}$ and covariance $\mathbf{\Sigma}_k \in \mathbb{R}^{3\times3}$ defined in world space.

To constrain $\mathbf{\Sigma}_k$ to the space of valid covariance matrices, a semi-definite parameterization $\mathbf{\Sigma}_k = \boldsymbol{O}_k \boldsymbol{s}_k \boldsymbol{s}_k^T \boldsymbol{O}_k^T$ is used. In this parameterization, $\boldsymbol{s}_k \in \mathbb{R}^3$ represents a scaling vector, while $\boldsymbol{O}_k \in \mathbb{R}^{3\times3}$ is a rotation matrix, defined through a quaternion.

To render an image for a given viewpoint, 3D Gaussian G_k is converted into camera coordinates and then projected into ray space to obtain the corresponding 2D Gaussian $\mathcal{G}_k^{2D}$ with a 2D covariance matrix $\mathbf{\Sigma}_k^{2D}$. Then, 3DGS (Kerbl et al. 2023) employs spherical harmonics to model view-dependent LDR color $\mathbf{c}_k^l$ and renders the image via alpha blending, which is applied in an order based on the depth of the 2D Gaussian (from 1 to N),

$$\mathbf{c}^l(\mathbf{x}) = \sum_{n=1}^{N} \mathbf{c}_n^l \alpha_n \mathcal{G}_n^{2D}(\mathbf{x}) \prod_{j=1}^{n-1} \left(1 - \alpha_j \mathcal{G}_j^{2D}(\mathbf{x})\right). \tag{2}$$

Given that the rendering process is both rapid and differentiable, it facilitates the efficient optimization of 3D Gaussian parameters through a multi-view loss, where 3D Gaussians are dynamically adjusted, being added or removed as needed to enhance scene representation. Readers could find more details in (Kerbl et al. 2023).

Event generation model. An event $e = (\mathbf{u}, t, p)$ at pixel position $\mathbf{u} = (u, v)$ and time t with polarity $p \in \{-1, +1\}$ is generated when the logarithmic change of brightness L since the last event at the pixel $\mathbf{x}$ and time $t - \Delta\mathrm{t}$ exceeds a threshold Θ ($\Theta > 0$) and the event e can be represented (Rebecq et al. 2019b) as:

$$e(\mathbf{u}, t) = \left\lfloor \frac{\log(\mathrm{L}(\mathbf{u}, \mathrm{t})) - \log(\mathrm{L}(\mathbf{u}, \mathrm{t} - \Delta\mathrm{t}))}{\Theta} \right\rfloor, \tag{3}$$

where the logarithmic change of brightness L represents the HDR radiance variations in the scene.

LDR image generation model. LDR image represents the integration of radiance L over the exposure time, which is then processed by the Camera Response Function (CRF):

$$I_{\mathrm{LDR}} = f_{\mathrm{CRF}}(I_{\mathrm{HDR}}), \tag{4}$$

where f_{CRF} is the function of the CRF, and the I_{HDR} can be expressed as:

$$I_{\mathrm{HDR}} = L \cdot \Delta\mathrm{t}, \tag{5}$$

where $\Delta\mathrm{t}$ is the expourse time.

Proposed EvHDR-GS

Formulation of HDR Video Reconstruction

Given a single-expourse LDR video $\mathcal{V}^l = \{\mathbf{I}_0^l, \mathbf{I}_1^l, ..., \mathbf{I}_n^l\}$ and events streams $\mathcal{E} = \{\mathbf{e}_0, \mathbf{e}_1, ..., \mathbf{e}_m\}$, our objective is to transform these LDR frames to HDR frames $\mathcal{V}^h = \{\mathbf{I}_0^h, \mathbf{I}_1^h, ..., \mathbf{I}_n^h\}$. Figure 2 shows the overview of our method.

Based on our observation, a video is essentially a projection of the 3D world onto the 2D image plane at different moments (or different viewpoints). Thus, the HDR video reconstruction task could be treated as the HDR scene reconstruction, and the HDR frame estimation is also transferred to HDR view rendering. To reconstruct an HDR scene, we first reproject each LDR frame in the video back to the 3D scene based on the camera information estimated by Structure from Motion (SfM) techniques (Schonberger and Frahm 2016). Subsequently, we perform HDR reconstruction by utilizing the event information during the reconstruction process. Once the optimization is finished, we could render the HDR frame from the HDR scene by the camera information corresponding to the frame index. The key advantage of this method lies in leveraging the multi-view consistency of the 3D scene to ensure the temporal consistency between successive frames in the video. This is because these frames are essentially projections of the 3D scene onto the 2D image plane at closely successive moments.

Thus, in our approach, HDR video reconstruction effectively utilizes LDR video frames $\mathcal{V}^l$ and the corresponding event streams $\mathcal{E}$ to reconstruct an HDR 3D scene. This reconstructed HDR 3D scene is then reprojected onto the 2D image plane at various moments, resulting in the production of the final HDR video frames $\mathcal{V}^h$. This method ensures a consistent rendering of HDR content across different viewing instances.

HDR 3D Gaussians

Since vanilla 3D Gaussians can only render LDR images in the range of (0, 255), We introduce HDR 3D Gaussians to represent HDR colors. Specifically, we additionally add an HDR branch to each 3D Gaussian to have both HDR color

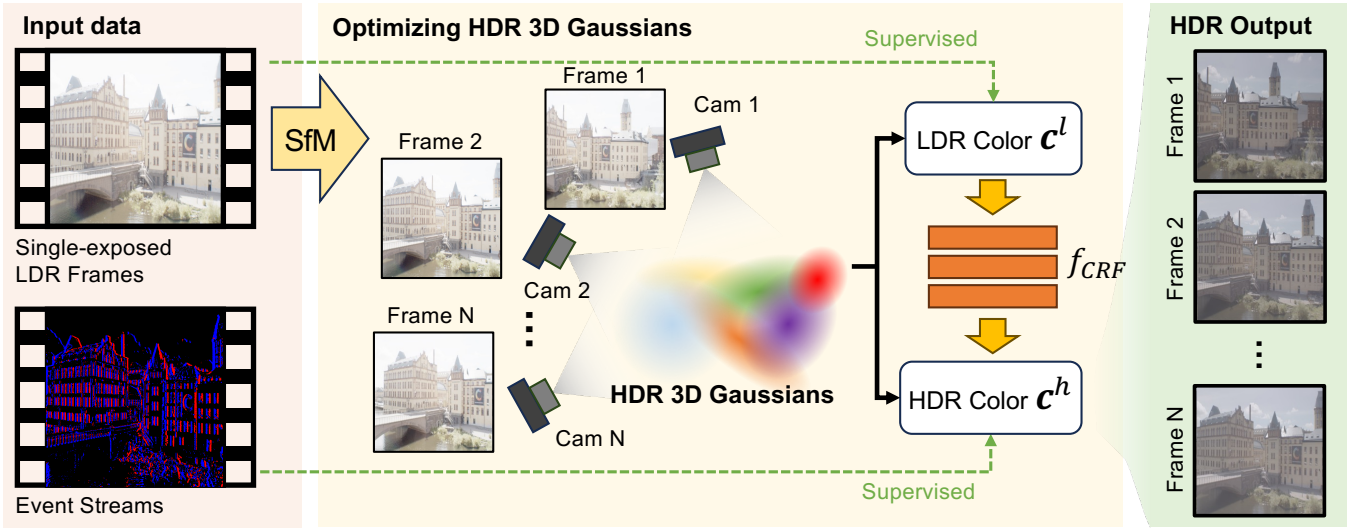

Figure 2: Overview of the proposed EvHDR-GS. We first utilize Structure from Motion (SfM) techniques (Schonberger and Frahm 2016) to calculate the camera poses for each frame in the video in the world space. Subsequently, the estimated camera poses and input single-exposed LDR frames are used for optimizing the HDR 3D Gaussians. We model the HDR-to-LDR transformation for HDR and LDR colors as CRF f_{CRF} and supervise it by paired event streams. After optimization, we utilize HDR 3D Gaussians to render all HDR frames at corresponding camera poses for HDR video reconstruction.

$\mathbf{c}^h$ and LDR color $\mathbf{c}^l$. Therefore, the pixel $\mathbf{c}^h(\mathbf{x})$ in the HDR image $\hat{\mathbf{I}}^h$ is rendered with the alpha blending of splatted 3D Gaussians on 2D image plane (similar to Equation (2)),

$$\mathbf{c}^h(\mathbf{x}) = \sum_{n=1}^{N} \mathbf{c}_n^h \alpha_n \mathcal{G}_n^{2D}(\mathbf{x}) \prod_{j=1}^{n-1} \left(1 - \alpha_j \mathcal{G}_j^{2D}(\mathbf{x})\right), \quad (6)$$

where $\mathbf{c}^h$ is represented by the logged spherical harmonics SH as $\mathbf{c}^h = \exp(\text{SH})$.

Therefore, the rendered HDR image $\hat{\mathbf{I}}^h$ represents the HDR frame corresponding to the ground truth LDR video frame $\mathbf{I}^l$, and we optimize HDR image $\hat{\mathbf{I}}^h$ with the estimated LDR image $\hat{\mathbf{I}}^l$ through following HDR-to-LDR transformation.

HDR-to-LDR transformation. We model the HDR-to-LDR transformation for the rendered HDR color $\mathbf{c}^h$ and LDR color $\mathbf{c}^l$,

$$\mathbf{c}^l = f_{\text{CRF}}(\mathbf{c}^h \cdot \Delta\text{t}), \quad (7)$$

where Δt is the exposure time and the tone-mapping operation $f_{\text{CRF}}(\cdot)$ models CRF that non-linearly maps the HDR color $\mathbf{c}^h$ into the LDR color $\mathbf{c}^l$.

Follow (Huang et al. 2022; Debevec and Malik 2023), we transform $f_{\text{CRF}}(\cdot)$ from linear domain to logarithmic domain to decouple the variables. The Equation (7) can be reformulated as:

$$\mathbf{c}^l = \mathcal{F}(\log(\mathbf{c}^{\text{h}}) + \log(\Delta\text{t})), \quad (8)$$

where $\mathcal{F} = (\ln f_{\text{CRF}}^{-1})^{-1}$.

Optimization of HDR 3D Gaussians

We then proceed to explain how our HDR 3D Gaussians is optimized by integrating both the event streams $\mathcal{E}$ and the LDR video frames $\mathcal{V}^l$, enhancing the dynamic range of the output video frames.

Events for HDR Reconstruction. Events capture the HDR information in the scene, but they are sparse and contain noise. We accumulate events between two frames I_i and I_j to reduce noise,

$$\mathbf{E}_{I_i \to I_j} = \Theta \sum_{k=i}^{j} e_k, \quad (9)$$

where e_k are events between $\mathbf{I}_i$ and $\mathbf{I}_j$. And Equation (3) can be written as:

$$\hat{\mathbf{E}}_{\text{I}_\text{i} \to \text{I}_\text{j}} = \log(\hat{\mathbf{I}}_\text{j}^\text{h}) - \log(\hat{\mathbf{I}}_\text{i}^\text{h}). \quad (10)$$

The overall loss function for events is combined with L1 loss $\mathcal{L}_1$ and D-SSIM loss $\mathcal{L}_{\text{D-SSIM}}$,

$$\begin{aligned} \mathcal{L}_{\text{event}} &= \mathcal{L}_1(\hat{\mathbf{E}} \odot \mathbf{M}_\text{c}, \mathbf{E} \odot \mathbf{M}_\text{c}) \\ &+ \lambda \cdot \mathcal{L}_{\text{D-SSIM}}(\hat{\mathbf{E}} \odot \mathbf{M}_\text{c}, \mathbf{E} \odot \mathbf{M}_\text{c}), \end{aligned} \quad (11)$$

where $\mathbf{M}_\text{c}$ is the color mask as described in (Rudnev et al. 2023), λ is a hyperparameter which is set as 0.2 across all evaluations. We found that incorporating a D-SSIM loss based on events can enhance the HDR reconstruction results.

Optimization of CRF. To stabilize the optimization of CRF, we model $\mathcal{F}$ with an exponential function $\mathcal{F}(x) = x^a$ parameterized by an exponential term a. We set the learning rate of this term a to 0.001.

Frames for HDR Reconstruction. Similar to the 3DGS method (Kerbl et al. 2023), we use LDR frames $\mathbf{I}^l$ from the input video $\mathcal{V}^l$ to supervised the rendered frame $\hat{\mathbf{I}}^l$ based on HDR 3D Gaussian model,

$$\mathcal{L}_{\text{ldr}} = \mathcal{L}_1(\hat{\mathbf{I}}^l, \mathbf{I}^l) + \lambda \cdot \mathcal{L}_{\text{D-SSIM}}(\hat{\mathbf{I}}^l, \mathbf{I}^l), \tag{12}$$

where λ is same as that used in event loss $\mathcal{L}_{\text{event}}$.

Therefore, the overall training loss is,

$$\mathcal{L} = \mathcal{L}_{\text{ldr}} + \mathcal{L}_{\text{event}}. \tag{13}$$

Implementation Details. We implement our method based on 3DGS (Kerbl et al. 2023) and add HDR outputs in the SH module for the original 3DGS. We adopt the same hyperparameter settings as described in 3DGS. All experiments are conducted on one single NVIDIA A5000.

Experiments

Experimental Setups

Compared methods. We compare the proposed EvHDR-GS to several state-of-the-art HDR imaging methods, including four frame-based HDR image reconstruction methods (Liu et al. 2020; Chung and Cho 2023; Xu et al. 2024; Cui et al. 2024a), a colored variant of an event-based HDR video reconstruction method (Rebecq et al. 2019b), and an event-guided HDR video reconstruction method (Yang et al. 2023). We conduct evaluations for other methods using the released codes and pre-trained weights provided in their papers. In the comparison with real-world scenes, since no device is available to capture multi-exposure videos, we only compare with frame-based methods that use single-exposure as input.

Synthetic data. Our synthetic data comes from two components: one part consists of static HDR videos from the DeepHDRVideo-Dataset (Chen et al. 2021a), and the other part is rendered using Blender files published by HDR-NeRF (Huang et al. 2022) and HDR-plenoxels (Jun-Seong et al. 2022). The synthetic data comprises 10 scenes, including indoor and outdoor environments, with 2,000 HDR frames. To generate events for these synthetic data, we implement the event generation method in (Yang et al. 2023). For input LDR images, we produce them based on the HDR-LDR conversion presented in (Huang et al. 2022). We follow (Huang et al. 2022; Jun-Seong et al. 2022) to use COLMAP (Schonberger and Frahm 2016) to obtain the camera pose of each frame.

Real-World Data. To validate the robustness of the proposed approach, we collected real-world data using the DAVIS 346C sensor under 4 scenes. In each scene, we capture the 6-second monocular event and image videos (each has 200 frames). We process the same camera calibration process as the synthetic data by using COLMAP.

Metric. We utilize standard metrics to evaluate HDR reconstruction results, including PSNR (Hore and Ziou 2010), SSIM (Hore and Ziou 2010), and LPIPS (Zhang et al. 2018). Consistent with common practice and following (Yan et al. 2019; Huang et al. 2022; Yang et al. 2023), we leverage the widely-used tone-mapping process μ-law for HDR reconstruction, which is expressed as,

$$M(E) = \frac{\log(1+\mu \text{E})}{\log(1+\mu)}, \tag{14}$$

where μ is the compression level set at 5000, and E is the scaled HDR pixel value within a range of [0, 1].

Besides, we also report the HDR-VDP-3 (Mantiuk et al. 2011) (referred to as 'VDP') metric for directly evaluating the quality of reconstructed HDR frames.

Method	PSNR↑	SSIM↑	LPIPS↓	VDP↑
Liu et al.	18.44	0.705	0.213	6.971
Rebecq et al.	10.86	0.698	0.346	4.264
Yang et al.	18.89	0.873	0.250	7.159
Chung and Cho	25.42	0.659	0.089	8.356
Xu et al.	25.09	0.624	0.104	8.276
Cui et al.	25.13	0.711	0.090	8.131
Ours w/o event	17.14	0.680	0.257	6.136
Ours	**27.94**	**0.971**	**0.037**	**8.651**

Table 1: Quantitative comparison on synthetic data. The best values are highlighted in **Bold**.

Effectiveness of HDR Video Reconstruction

Evaluation on synthetic data. We report quantitative results in Table 1. It is evident from these results that our method outperforms the counterparts with a distinguished improvement. Specifically, our EvHDR-GS achieves the best SSIM and LPIPS, which indicates that the reconstructed HDR frames preserve the structure and semantic information from input LDR frames, while other methods are likely to degrade both information. Besides, as the VDP metric is designed for direct HDR evaluation with the perceptual process, the best results on VDP validate the effectiveness of our method.

Figure 3 demonstrates the visual quality comparisons. We observe that our HDR reconstruction results are the closest to the ground truth (GT), which is consistent with the quantitative results. From the two middle outdoor scenes, it can be observed that Yang et al. exhibits a yellowish tint. This color bias stems from their method being trained on predefined events and frame data, causing the output distribution to closely resemble that of the training data, which leads to a yellow-tinged final result. Observing the overexposed regions in Figure 3, it can be seen that the event-guided methods have successfully restored the details in these areas, while the frame-based methods fail. The reason might be that frame-based methods can only attempt to reconstruct details based on prior assumptions in overexposed regions.

Evaluation on real-world data. We further evaluate the proposed method on real-world scenarios and demonstrate the qualitative results in Figure 4. We could observe that our method is capable of reconstructing HDR frames that exhibit a wider dynamic range and enhanced detail. Yang et al. produces results with a yellowish tint due to the data

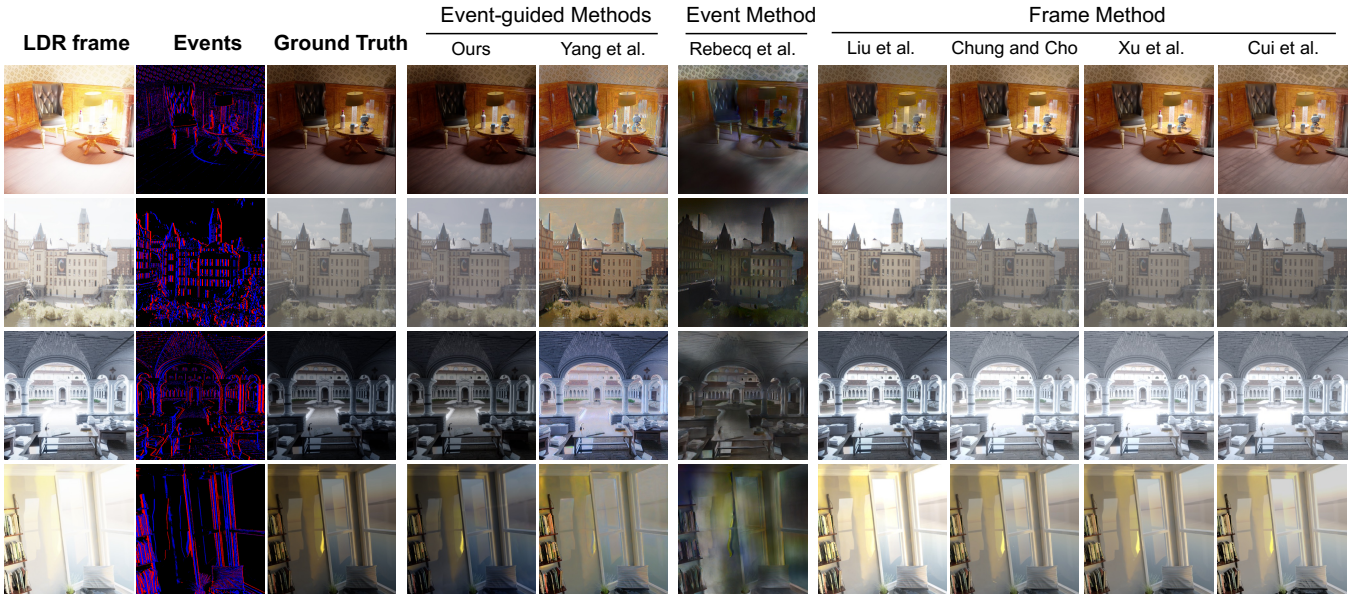

Figure 3: Qualitative comparisons on synthetic data. The 1st, 2nd, and 3rd columns are input LDR frames, input events, and ground truth HDR frames, respectively. The 4th to last columns are the reconstructed HDR frames by our and other methods. Note that HDR images are visualized by applying tone mapping using the μ-law. Zoom in for better details.

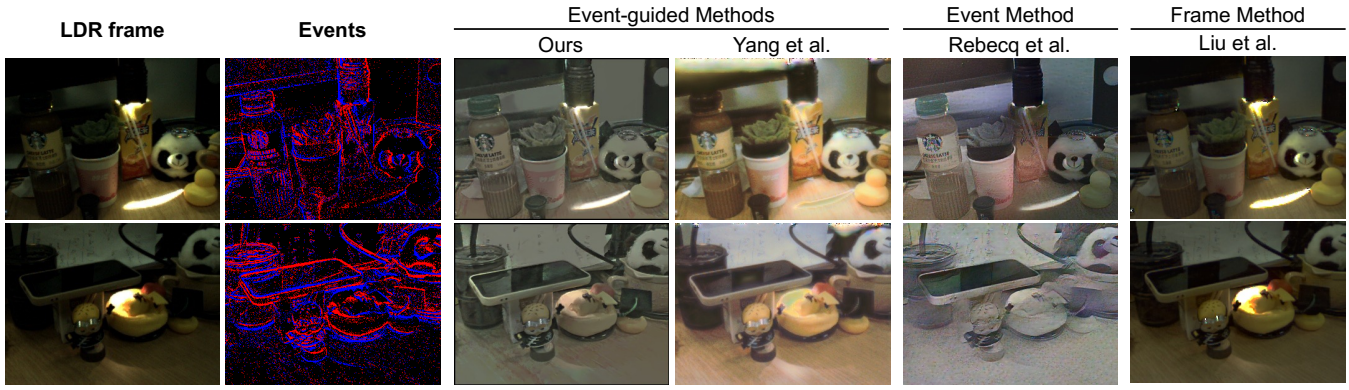

Figure 4: Qualitative comparisons on real-world data. The 1st, 2nd, and 3rd columns are input LDR frames, input events, and ground truth HDR frames, respectively. The 4th to last columns are the reconstructed HDR frames by our and other methods. Note that HDR images are visualized by applying tone mapping using the μ-law. Zoom in for better details.

bias through the training procedure. In contrast, our method directly reconstructs HDR video from the existing data distribution, resulting in outputs that more faithfully represent real-world scenes.

These results from both synthetic and real-world scenarios validate the effectiveness of our methods in reconstructing a better and consistent HDR video.

Effectiveness of Temporal Consistency

Figure 5 shows the visualization of temporal consistency. We conduct the evaluation on the temporal consistency to validate the effectiveness of temporal consistency brought by the 3D scene representation. Figure 5 demonstrates the adjacent HDR frames reconstructed by different methods and our EvHDR-GS. Compared to the event-guided HDR video reconstruction method (Yang et al. 2023) (Figure 5(A)), our method maintains the same color across these frames, while the counterpart shows color changes, especially in the boxed area. This is because the 3D consistency of the proposed HDR scene representation imposes the temporal consistency of the reconstructed HDR frames. On the other hand, in the comparison with the frame-based HDR reconstitution method (Figure 5(B)), we observe that Liu et al. displays different brightness levels in each frame, especially at the position of the desk lamp. The reason might be that the frame-based method lacks physical information to handle a consistent brightness on the overexposed region. In contrast, our method utilizes event information, allowing the overex-

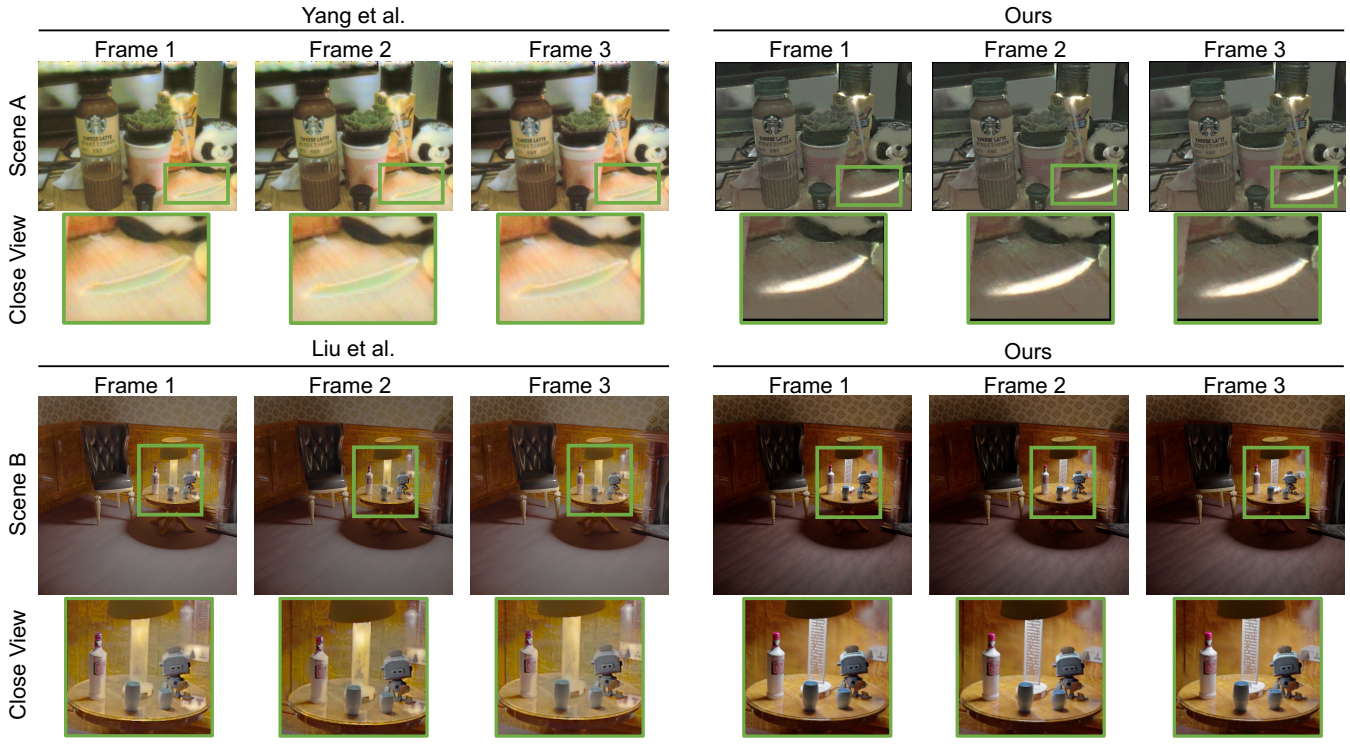

Figure 5: Qualitative comparisons on temporal consistency.(A) comparison with the event-guided HDR video reconstruction method (Yang et al. 2023); (B) comparison with the frame-based HDR reconstruction method (Liu et al. 2020). Note that HDR images are visualized by applying tone mapping using the μ-law.

Method	Time
Liu et al.	230ms
Rebecq et al.	63.56ms
Yang et al.	5278ms
Chung and Cho	132ms
Ours	7min (Optimization) + 10.4ms (Render)

Table 2: Inference time comparison with existing methods. We tested our approach using a 100-frame video with a resolution of 512x512 on an A5000.

posed area to exhibit the same brightness across frames.

These results validate the proposed HDR scene representation could significantly improve the temporal consistency of the reconstructed HDR video.

Validation on Event Information

To validate the effectiveness of the auxiliary events, we compare the proposed method with its variant without event loss $\mathcal{L}_{\text{event}}$, namely "Ours w/o event". Table 1 presents the quantitative comparison, where we observe that our performance degrades dramatically when removing event loss $\mathcal{L}_{\text{event}}$, which is worse than most of the existing HDR reconstruction methods. The reason is that the event information is crucial for HDR information reconstruction, as the single-exposed LDR frames could not provide the HDR information.

Conlcusion

In this paper, we present an event-guided HDR video reconstruction method based on 3D Gaussian Splatting. Based on our observation, a video is essentially a projection of the 3D world onto the 2D image plane at different moments. Thus, the HDR video reconstruction task could be treated as the HDR scene reconstruction. We use 3D Gaussians to represent the entire video, ensuring consistent brightness in the HDR output. Since existing 3D Gaussians are limited to representing the low dynamic range color (0-255), we have designed an HDR 3D Gaussian that can represent both HDR color and LDR color simultaneously. Additionally, we employ learnable parameters to model the relationship between events and LDR images. Our method is self-supervised, eliminating data bias present in existing event-guided HDR video reconstruction approaches. The proposed method demonstrates state-of-the-art performance both quantitatively and qualitatively for synthetic and real-world data.

Limitations. Our current work is limited to static scenarios as 3DGS could only represent the static scene. The enhanced method handles dynamic scenarios by replacing the vanilla 3D Gaussian Splitting with dynamic 3D Gaussian Splitting methods (Yang et al. 2024; Lin et al. 2024; Wu et al. 2024; Huang et al. 2024; Sun et al. 2024). Additionally, employing temporal encoding instead of positional encoding presents another way to handle dynamic scenarios.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61925603, 62376247, U20A20220, and 62334014), in part by the grant from Key R&D Program of Zhejiang (2022C01048), and in part by the Fundamental Research Funds for the Central Universities.

References

Bardow, P.; Davison, A. J.; and Leutenegger, S. 2016. Simultaneous optical flow and intensity estimation from an event camera. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Chen, G.; Chen, C.; Guo, S.; Liang, Z.; Wong, K.-Y. K.; and Zhang, L. 2021a. HDR Video Reconstruction: A Coarse-to-fine Network and A Real-world Benchmark Dataset.

Chen, Z.; Lu, Z.; Ma, D.; Tang, H.; Jiang, X.; Zheng, Q.; and Pan, G. 2024. Event-ID: Intrinsic Decomposition Using an Event Camera. In *Proceedings of the 32nd ACM International Conference on Multimedia*.

Chen, Z.; Zheng, Q.; Niu, P.; Tang, H.; and Pan, G. 2021b. Indoor lighting estimation using an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Chung, H.; and Cho, N. I. 2023. Lan-hdr: Luminance-based alignment network for high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Cui, J.; Jiang, W.; Peng, Z.; Pan, Z.; and Cao, Z. 2024a. Exposure Completing for Temporally Consistent Neural High Dynamic Range Video Rendering. In *Proceedings of the 32nd ACM International Conference on Multimedia*.

Cui, M.; Wang, Z.; Wang, D.; Zhao, B.; and Li, X. 2024b. Color Event Enhanced Single-Exposure HDR Imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Debevec, P. E.; and Malik, J. 2023. Recovering high dynamic range radiance maps from photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*.

Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R. K.; and Unger, J. 2017. HDR image reconstruction from a single exposure using deep CNNs. *ACM transactions on graphics*.

Endo, Y.; Kanamori, Y.; and Mitani, J. 2017. Deep reverse tone mapping. *ACM Trans. Graph*.

Han, J.; Yang, Y.; Duan, P.; Zhou, C.; Ma, L.; Xu, C.; Huang, T.; Sato, I.; and Shi, B. 2023. Hybrid high dynamic range imaging fusing neuromorphic and conventional images. *IEEE Transactions on pattern analysis and machine intelligence*.

Han, J.; Zhou, C.; Duan, P.; Tang, Y.; Xu, C.; Xu, C.; Huang, T.; and Shi, B. 2020. Neuromorphic camera guided high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Hore, A.; and Ziou, D. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*. IEEE.

Huang, X.; Zhang, Q.; Feng, Y.; Li, H.; Wang, X.; and Wang, Q. 2022. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Huang, Y.-H.; Sun, Y.-T.; Yang, Z.; Lyu, X.; Cao, Y.-P.; and Qi, X. 2024. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Jun-Seong, K.; Yu-Ji, K.; Ye-Bin, M.; and Oh, T.-H. 2022. Hdr-plenoxels: Self-calibrating high dynamic range radiance fields. In *European Conference on Computer Vision*.

Kalantari, N. K.; and Ramamoorthi, R. 2019. Deep HDR video from sequences with alternating exposures. In *Computer graphics forum*. Wiley Online Library.

Kang, S. B.; Uyttendaele, M.; Winder, S.; and Szeliski, R. 2003. High dynamic range video. *ACM Trans On Graphics*.

Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.*

Kim, H.; Handa, A.; Benosman, R.; Ieng, S.-H.; and Davison, A. J. 2008. Simultaneous mosaicing and tracking with an event camera. *J. Solid State Circ*.

Lele, A.; Fang, Y.; Ting, J.; and Raychowdhury, A. 2021. An end-to-end spiking neural network platform for edge robotics: From event-cameras to central pattern generation. *IEEE Trans on Cognitive and Developmental Systems*.

Li, H.; Li, G.; Ji, X.; and Shi, L. 2018. Deep representation via convolutional neural network for classification of spatiotemporal event streams. *Neurocomputing*.

Li, X.; Lu, Q.; Fan, C.; Zhao, C.; Zou, L.; and Yu, L. 2024. Generalizing event-based HDR imaging to various exposures. *Neurocomputing*.

Liang, Z.; Zhang, Q.; Feng, Y.; Shan, Y.; and Jia, K. 2024. Gs-ir: 3d gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Liao, Z.; Liu, Y.; Zheng, Q.; and Pan, G. 2024. Spiking NeRF: Representing the Real-World Geometry by a Discontinuous Representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 13790–13798.

Lin, Y.; Dai, Z.; Zhu, S.; and Yao, Y. 2024. Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Liu, Y.-L.; Lai, W.-S.; Chen, Y.-S.; Kao, Y.-L.; Yang, M.-H.; Chuang, Y.-Y.; and Huang, J.-B. 2020. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Liu, Z.; Wang, Y.; Zeng, B.; and Liu, S. 2022. Ghost-free high dynamic range imaging with context-aware transformer. In *European Conference on computer vision*.

Lu, Z.; Zheng, Q.; Shi, B.; and Jiang, X. 2024. Pano-NeRF: Synthesizing High Dynamic Range Novel Views with Geometry from Sparse Low Dynamic Range Panoramic Images. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Mantiuk, R.; Kim, K. J.; Rempel, A. G.; and Heidrich, W. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*.

Messikommer, N.; Georgoulis, S.; Gehrig, D.; Tulyakov, S.; Erbach, J.; Bochicchio, A.; Li, Y.; and Scaramuzza, D. 2022. Multi-bracket high dynamic range imaging with event cameras. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*.

Mostafavi, M.; Wang, L.; and Yoon, K.-J. 2021. Learning to reconstruct hdr images from events, with applications to depth and flow prediction. *International Journal of Computer Vision*.

Nayar, S. K.; and Mitsunaga, T. 2000. High dynamic range imaging: Spatially varying pixel exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*. IEEE.

Pan, L.; Hartley, R.; Scheerlinck, C.; Liu, M.; Yu, X.; and Dai, Y. 2020. High frame rate video reconstruction based on an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Pan, L.; Scheerlinck, C.; Yu, X.; Hartley, R.; Liu, M.; and Dai, Y. 2019. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Rebecq, H.; Ranftl, R.; Koltun, V.; and Scaramuzza, D. 2019a. Events-to-video: Bringing modern computer vision to event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Rebecq, H.; Ranftl, R.; Koltun, V.; and Scaramuzza, D. 2019b. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*.

Rempel, A. G.; Trentacoste, M.; Seetzen, H.; Young, H. D.; Heidrich, W.; Whitehead, L.; and Ward, G. 2007. Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM transactions on graphics (TOG)*.

Rudnev, V.; Elgharib, M.; Theobalt, C.; and Golyanik, V. 2023. Eventnerf: Neural radiance fields from a single colour event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Samra, R.; Mitra, K.; and Shedligeri, P. 2023. High-Speed HDR Video Reconstruction from Hybrid Intensity Frames and Events. In *Computer Vision and Machine Intelligence: Proceedings of CVMI 2022*.

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

Shu, Y.; Shen, L.; Hu, X.; Li, M.; and Zhou, Z. 2024. Towards Real-World HDR Video Reconstruction: A Large-Scale Benchmark Dataset and A Two-Stage Alignment Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Song, J. W.; Park, Y.-I.; Kong, K.; Kwak, J.; and Kang, S.-J. 2022. Selective transhdr: Transformer-based selective hdr imaging using ghost region mask. In *European Conference on Computer Vision*.

Sun, J.; Jiao, H.; Li, G.; Zhang, Z.; Zhao, L.; and Xing, W. 2024. 3dgstream: On-the-fly training of 3d gaussians for efficient streaming of photo-realistic free-viewpoint videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Tocci, M. D.; Kiser, C.; Tocci, N.; and Sen, P. 2011. A versatile HDR video production system. *ACM Trans on Graphics*.

Wang, L.; Ho, Y.-S.; Yoon, K.-J.; et al. 2019. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Wang, L.; Kim, T.-K.; and Yoon, K.-J. 2020. Eventsr: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Wu, G.; Yi, T.; Fang, J.; Xie, L.; Zhang, X.; Wei, W.; Liu, W.; Tian, Q.; and Wang, X. 2024. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Xu, G.; Wang, Y.; Gu, J.; Xue, T.; and Yang, X. 2024. HDR-Flow: Real-Time HDR Video Reconstruction with Large Motions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yan, Q.; Gong, D.; Shi, Q.; Hengel, A. v. d.; Shen, C.; Reid, I.; and Zhang, Y. 2019. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yang, Y.; Han, J.; Liang, J.; Sato, I.; and Shi, B. 2023. Learning event guided high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yang, Z.; Gao, X.; Zhou, W.; Jiao, S.; Zhang, Y.; and Jin, X. 2024. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Zhang, W.; Li, Z.; Ma, D.; Tang, H.; Jiang, X.; Zheng, Q.; and Pan, G. 2024. Spiking GS: Towards High-Accuracy and Low-Cost Surface Reconstruction via Spiking Neuron-based Gaussian Splatting. *arXiv preprint arXiv:2410.07266*.

Zou, Y.; Zheng, Y.; Takatani, T.; and Fu, Y. 2021. Learning to reconstruct high speed and high dynamic range videos from events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024–2033.