An Event Camera Simulator for Arbitrary Viewpoints based on Neural Radiance Fields

Diego Hernández Rodríguez[1,2], Motoharu Sonogashira[1,2], Kazuya Kitano[2], Yuki Fujimura[2], Takuya Funatomi[2][a], Yasuhiro Mukaigawa[2], and Yasutomo Kawanishi[1,2][b]

[1] *Guardian Robot Project, RIKEN, Kyoto, JAPAN*

[2] *Division of Information Science, Nara Institute of Science and Technology, Nara, JAPAN*
hernandez_rodriguez.diego.hc7@is.naist.jp, yasutomo.kawanishi@riken.jp

Keywords: Event camera simulation, Neural radiance fields

Abstract: Event cameras are novel sensors that offer significant advantages over standard cameras, such as high temporal resolution, high dynamic range, and low latency. Despite recent efforts, however, event cameras remain relatively expensive and difficult to obtain. Simulators for these sensors are crucial for developing new algorithms and mitigating accessibility issues. However, existing simulators based on a real-world video often fail to generalize to novel viewpoints or temporal resolutions, making the generation of realistic event data from a single scene unfeasible. To address these challenges, we propose enhancing event camera simulators with neural radiance fields (NeRFs). NeRFs can synthesize novel views of complex scenes from a low-frame-rate video sequence, providing a powerful tool for simulating event cameras from arbitrary viewpoints. This approach not only simplifies the simulation process but also allows for greater flexibility and realism in generating event camera data, making the technology more accessible to researchers and developers.

1 INTRODUCTION

Event cameras represent a paradigm shift in visual sensing technology, capturing dynamic scenes with remarkable temporal resolution and high dynamic range. Unlike conventional frame-based cameras, event cameras asynchronously record changes in the intensity of the visual field, offering a unique advantage in scenarios involving fast motion or challenging lighting conditions. Since these sensors are still relatively expensive and difficult to obtain, various efforts have been made to create simulators to facilitate their research further.

Previous simulators aim to generate event data from RGB video by either relying on ultra-high framerates (Gehrig et al., 2020; García et al., 2016) or by interpolation of the video sequence (Hu et al., 2021). This comes with the drawback of not being able to generate more data from a single video. While simulators like ESIM (Rebecq et al., 2018) attempt to tackle this issue with the use of 3D models, generating data that resembles a realistic scene is both time and labor-intensive, making it unsuitable for researchers who want to generate for their own environment.

To address this challenge of generating event data from arbitrary viewpoints, we propose a framework of a simulation shown in Fig. 2. The framework generates synthetic event camera data using neural radiance fields (NeRFs) (Mildenhall et al., 2020), a recent breakthrough in the field of computer vision that enables the reconstruction of high-fidelity 3D scenes from a sparse set of 2D images by leveraging neural networks to model the volumetric radiance field. By integrating NeRF with event-based sensing principles, we aim to create a versatile framework that can produce realistic and diverse event camera data, facilitating the advancement of event-based vision algorithms. Notably, our method focuses on generating event data from static scenes, allowing for the exploration of how camera motion alone influences event generation without the added complexity of dynamic scene changes.

Our approach offers several significant advantages. First, it allows for creating extensive datasets without the need for labor-intensive data collection processes. Second, it provides a controlled virtual environment where various parameters can be modified to evaluate the robustness of event-based algorithms. Finally, the synthetic data generated through

[a] https://orcid.org/0000-0001-5588-5932

[b] https://orcid.org/0000-0002-3799-4550

our method can serve as a valuable resource for training deep learning models, potentially improving their performance in real-world applications.

The rest of this paper is organized as follows. In section 2, we introduce some of the most important works concerning event camera simulation and explain their working mechanism. We also quickly review the formulation of neural radiance fields. In section 3, we detail the proposed methodology for synthesizing event camera data using NeRF and discuss the implementation and integration of these technologies. In section 4, we present experimental results demonstrating the effectiveness of our approach, comparing them to actual event data streams and with other video-to-event generation pipelines. Finally, in section 5, we discuss our method's limitations, possible extensions, and future work. By bridging the gap between synthetic data generation and event-based sensing, our work aims to accelerate research in event cameras and pave the way for their broader adoption and application.

2 RELATED WORK

2.1 Event camera simulation

Numerous event camera datasets and simulators have been introduced over the years. In this section, we review the most relevant ones and their specific application scenarios. The number of publicly released event camera simulators is small. While some of them build upon previous research, they mostly tackle the task differently.

Early simulators like (Mueggler et al., 2017) visually approximated an event stream by detecting significant changes in luminance between two successive frames to create edge-like images that resemble the output of an event camera. Most of these simulators did not discuss how to convert the simulated events into realistic and accurate raw event streams. Recent approaches like (Zhang et al., 2024) tackle this by developing a statistics-based local-dynamics-aware timestamp inference algorithm that enables the smooth transition to the event stream. Other simulators like (Joubert et al., 2021) attempt to physically model the unique characteristics of the sensor and its parameters, while methods such as (Zhu et al., 2019) and (Hu et al., 2021) take a deep learning-based approach in order to approximate the outputs of a physical sensor. However, none of them take into account the geometry of the scene, nor can they generate an event stream outside the original path followed by the camera. In order to circumvent this limitation, (Li et al., 2018) and, most notably, (Rebecq et al., 2018) leverage 3D models to render a scene in which a user-defined camera path can be utilized to generate an event stream. However, this approach poses the need for detailed models in case a realistic scene is to be simulated.

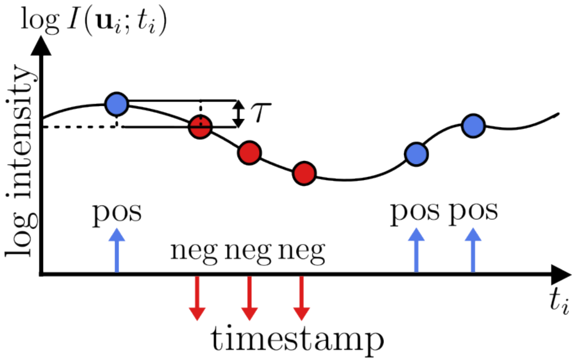


Figure 1: A pixel u of the intensity image I_t in the event generation model. A positive or negative event is generated when the brightness change exceeds the threshold τ in a logarithmic scale. Represented in blue and red, respectively.

In ESIM (Rebecq et al., 2018), an output event stream E is represented as a sequence of $\mathbf{e}_i = (t_i, \mathbf{u}_i, p_i)$, denoting brightness changes asynchronously registered by an image I at time t and its pixel location $\mathbf{u}_i = (x_i, y_i)$ in the image, with a polarity $p_i \in \{-1, 1\}$. The polarity of an event indicates a positive or negative change in illumination according to a logarithmic scale, quantized by negative and positive thresholds τ. The change in brightness between two timestamps can be estimated by the difference of intensity of a pixel $\mathbf{u}_i$ of images at time t_i and t_{i-1} in the logarithmic scale. This mechanism is illustrated in Fig. 1 and formulated as follows.

$$p_i = \begin{cases} -1 & \text{if} \quad \tau < \Delta(\mathbf{u}_i; t_i) \\ 1 & \text{if} \quad \tau > \Delta(\mathbf{u}_i; t_i) \end{cases} \tag{1}$$

$$\Delta(\mathbf{u}_i; t_i) = \log I(\mathbf{u}_i; t_i) - \log I(\mathbf{u}_i; t_{i-1}) \tag{2}$$

2.2 Neural radiance fields

Neural radiance fields (Mildenhall et al., 2020) represent a scene utilizing a multi-layer perceptron (MLP) $F_\theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$ that maps a position in 3D space $\mathbf{x} = (x, y, z)$ and a 2D viewing direction $\mathbf{d} = (\theta, \phi)$ to its corresponding directional emitted radiance, i.e., its color $\mathbf{c} = (\mathrm{R}, \mathrm{G}, \mathrm{B})$ and volume density σ. From this representation, the estimated emitted radiance $\widehat{\mathbf{L}}$ at a given pixel $\mathbf{u}$ can be calculated using the volume rendering equation (Tagliasacchi and Mildenhall, 2022)

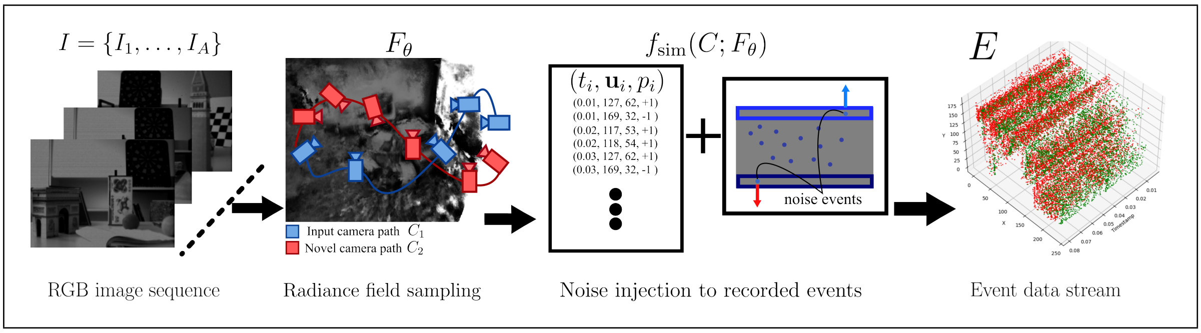


Figure 2: Illustration of our method. We first train a neural radiance field and subsequently simulate a virtual event camera, adding noise to the simulation

with quadrature, as follows:

$$\widehat{\mathbf{L}}(\mathbf{u}) = \sum_{k=1}^{N} T_k(1 - \exp(-\sigma_k \delta_k))\mathbf{c}_k, \tag{3}$$

$$T_k = \exp\left(-\sum_{m=1}^{k-1} \sigma_m \delta_m\right), \tag{4}$$

where σ_k and $\mathbf{c}_k$ are the volume density and the emitted radiance, respectively, of a sampled position $\mathbf{x}_k$ along the back-projected ray $\mathbf{r}$ through a pixel, which has a direction $\mathbf{d}$ and an origin $\mathbf{o}$ at the camera center. The sample $\mathbf{x}_k = \mathbf{o} + s_k\mathbf{d}$ has a distance s_k from the camera center and a distance of $\delta_k = s_{k+1} - s_k$ between its adjacent sample $\mathbf{x}_{k+1}$.

Several advances have been made since the original NeRF paper was first published. Neural network-based approaches like (Müller et al., 2022) and (Chen et al., 2022) have greatly reduced inference time and increased 3D reconstruction quality, while methods such as (Kerbl et al., 2023) completely forgo a neural representation and opt for a modified differentiable point-based rendering technique. While we utilize (Müller et al., 2022) as our rendering backbone in this paper, it is worth noting that our method is radiance field agnostic. Meaning that the method used to render the radiance field is interchangeable.

2.3 Event Cameras and NeRFs

Recent studies have explored the integration of event cameras with NeRFs. Notable works such as (Klenk et al., 2023), (Rudnev et al., 2023), and (Hwang et al., 2023) have demonstrated promising results in constructing radiance field representations directly from event camera data streams. In contrast, this paper shifts focus from generating radiance fields to deriving event representations from existing radiance fields.

This approach presents several advantages. By leveraging the continuous and high-resolution nature of NeRFs, it becomes possible to simulate event data from arbitrary viewpoints and under varying conditions without the need for specialized hardware. This flexibility enables the creation of diverse datasets for training and evaluating event-based algorithms, which are often limited by the scarcity and cost of event cameras.

However, simulating realistic event data from NeRFs introduces unique challenges. Reproducing sensor-specific noise and latency effects is essential for generating data that closely mirrors real-world conditions and addressing these challenges is critical to ensuring that the simulated events are both physically plausible and useful for downstream tasks.

3 METHOD

3.1 Problem formulation

Event cameras asynchronously detect changes in pixel brightness, delivering high temporal resolution and low-latency data. However, their high cost and limited availability restrict widespread adoption. Researchers rely on simulators to generate synthetic event streams, yet existing simulators that rely on RGB videos are limited to the viewpoints present in the input sequence, preventing the generation of event data from novel camera paths.

To address these challenges, we propose a NeRF-based event camera simulator capable of generating synthetic event data from arbitrary viewpoints.

The simulation consists of two stages: NeRF training and event data generation. In the NeRF training stage, the simulator accepts a set of images $I = \{I_1, \ldots, I_A\}$. Using I, a neural radiance field F_θ is trained. In the event data generation stage, a camera trajectory (a sequence of camera positions and orientations) $C = [\mathbf{c}_1, \ldots, \mathbf{c}_B]$ are input to the simulator f_{sim}. Then, the simulator generates an event data stream E

along the given camera path C using the trained neural radiance field F_θ.

$$E = f_{\text{sim}}(C; F_\theta) \tag{5}$$

The overall process flow is illustrated in Fig. 2.

3.2 Event data generation by sampling radiance fields

Following the methodology behind ESIM's event generation from 3D models, our method approximates the per-pixel value of the intensity image $\log I(\mathbf{u}_i; t_i)$ at pixel $\mathbf{u}_i$ by using the trained F_θ and a selected camera position interpolated from the given camera path C. For each pixel of the image at the sampled camera position, the color of each pixel is calculated by accumulating the contributions from all sampled points along the ray following equation (3). Since event cameras operate in brightness pixels, we convert the sampled color images using the *ITU-R Recommendation BT601* for luma (Union, 2011), i.e., according to the formula:

$$Y(R, G, B) = 0.299R + 0.587G + 0.114B, \tag{6}$$

with RGB channels in linear color space. This yields the following equation:

$$I(\mathbf{u}_i; t_i) = Y(\widehat{\mathbf{L}}(\mathbf{u}_i; t_i)). \tag{7}$$

Generating a pair of logarithmic intensity images $\log(I(\mathbf{u}_i; t_i))$ and $\log(I(\mathbf{u}_i; t_{i-1}))$ based on user-defined parameters, such as maximum number of events per camera position, pixel refraction period, and brightness change threshold τ.

We can then determine the number of predicted events at a certain pixel location during that time window with the following equation:

$$n_i = \left\lfloor \frac{|\Delta(\mathbf{u}_i; t)|}{\tau} \right\rfloor. \tag{8}$$

According to the number, events are generated with polarity p_i based on the positive or negative of $\Delta(\mathbf{u}_i; t)$.

3.3 Event camera noise

Although less studied than traditional RGB camera noise, a data stream from an event camera normally contains events that are not associated with changes in intensity. These events are considered noise which comes from two main sources: photon noise and leakage current (Guo and Delbruck, 2023). In low-brightness conditions, photon noise is the most common source of noise, while leakage current dominates high-brightness conditions. In some event camera simulators like (Hu et al., 2021), the events that are generated by photon noise are modeled as a Poisson process, in which the noise event rate linearly decreases with intensity. Further research on modeling these noise sources were performed by Ruiming et. al. (Cao et al., 2024) and we leverage their noise model in our experiments.

4 EVALUATION

4.1 Experimental settings

We utilize a modified version of instant NGP (Müller et al., 2022) implemented in PyTorch as our NeRF backbone. Each scene of the dataset was trained for 350 epochs with an initial learning rate of 0.01 and with the Adam optimizer. We conduct our experiments on the dataset provided by Mueggler et al. (Mueggler et al., 2017) for our comparisons since it contains images generated by a DAVIS sensor (Brandli et al., 2014), which are used to train the radiance field, as well as camera positions from an external tracker, eliminating the need to use COLMAP (Schönberger and Frahm, 2016) for camera pose estimation.

To perform our tests, we interpolate five equidistant positions between each camera pose along the initial camera path, akin to the frame interpolation V2E does.

4.2 Evaluation metrics

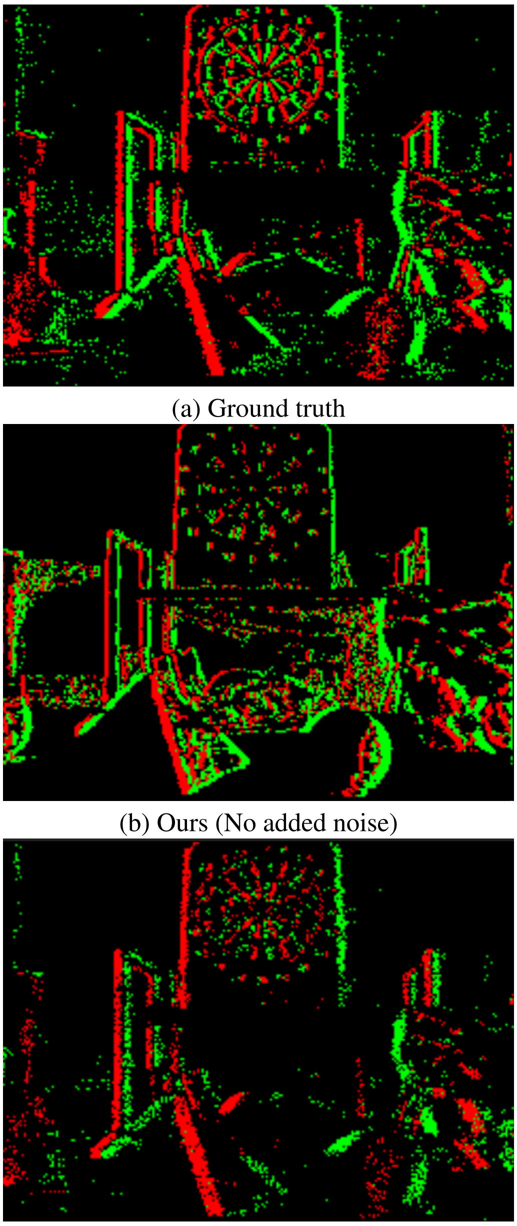
For evaluation, it is difficult to directly compare the generated event data and the ground truth, we accumulated events into an image and performed image-level comparison.

First, an accumulation operation is performed on both the ground truth and simulated event streams to generate a frame representation. The accumulation operation integrates events over time into a frame-by-frame basis, aggregating changes captured by the sensor. As shown in (Mueggler et al., 2017), a logarithmic intensity image $\log \hat{I}(\mathbf{u}; t)$ can be reconstructed from the event stream at any point in time t by accumulating events $\mathbf{e}_i = (t_i, \mathbf{u}_i, p_i)$ according to the following function:

$$\log \hat{I}(\mathbf{u}; t) = \log I(\mathbf{u}; 0) + \gamma(\mathbf{u}; t), \tag{9}$$

$$\gamma(\mathbf{u}; t) = \sum_{0 < t_i \leq t} p_i \tau \delta(\mathbf{u} - \mathbf{u}_i) \delta(t - t_i), \tag{10}$$

where $I(\mathbf{u}; 0)$ is the rendered image at time $t = 0$, and δ selects the pixel to be updated on every event (pixel $\mathbf{u}_i$ of $\hat{I}$ is updated at time t_i).

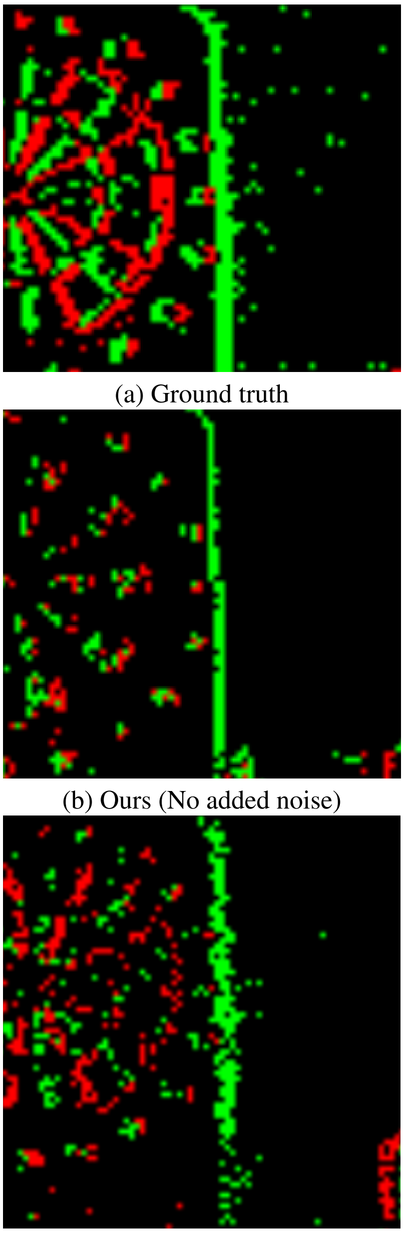


(a) Ground truth

(b) Ours (No added noise)

(c) Ours (with added noise)

Figure 3: Comparison of event streams, positive and negative events are colored red and green, respectively.



(a) Ground truth

(b) Ours (No added noise)

(c) Ours (with added noise)

Figure 4: Zoomed-in view of a specific region from Figure 3, highlighting finer details of the event streams.

We utilize a modified version of this function, which applies a decay parameter to reduce the noise of the generated frame. The accumulator function applies an exponential decay $d(t,\tau)$ to equation (7):

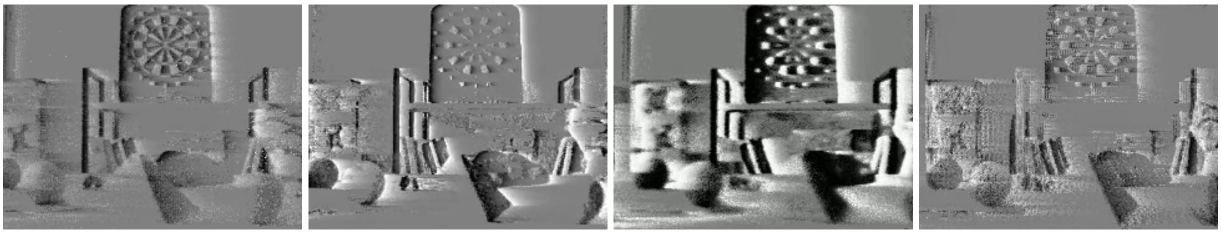
$$\begin{aligned}\log\widehat{I}(\mathbf{u};t) = &\log\Big(I(\mathbf{u};0)d(t,\tau) + I(\mathbf{u};t)(1-d(t,\tau)) \\ &+ \sum_{0<t_i\le t} p_i\tau\delta(\mathbf{u}-\mathbf{u}_i)d(t-t_k,\tau)\Big),\end{aligned} \tag{11}$$

$$d(t,\tau) = \exp\left(-\frac{t}{\tau}\right), \tag{12}$$

where $\log(I(\mathbf{u};0))$ is the logarithm of the intensity of the pixel at the previous accumulated frame, $\log(I(\mathbf{u};t))$ is a neutral potential, and the decay parameter is the time constant τ. For our experiments we set $\tau = 1\times10^{-5}$ microseconds and $\log(I(\mathbf{u};0))$ = 0.5.

4.2.1 Qualitative comparison

We compare frame-level results with the real event camera stream (ground truth). We also compare the accumulated results with V2E and V2CE to the real



(a) Ground truth (b) Ours (c) V2E (Hu et al., 2021) (d) V2CE (Zhang et al., 2024)

Figure 5: Visual comparison of accumulated frames. All frames were obtained after accumulating events according to the process described in section 4.2

Table 1: Comparison of PSNR (dB) values obtained in scenes from the dataset (Mueggler et al., 2017) (higher is better).

Scene name	Ours	V2E	V2CE
slider	**30.01**	29.40	29.42
boxes 6DoF	**28.32**	28.06	28.15
poster	28.04	28.57	**28.64**

event camera stream (accumulated ground truth).

4.2.2 PSNR of accumulated event frames

To measure the correctness of the simulated events quantitatively, we perform an evaluation using a Peak Signal Noise Ratio (PSNR) basis, which is a well-known evaluation metric of image quality (Horé and Ziou, 2010). These PSNR comparisons are summarized in Tab. 1.

4.3 Experimental results

As demonstrated in Fig. 3, our simulator correctly approximates the positive and negative events measured by an actual event camera. It is worth noting that due to not including both noise and hot pixel simulation in our experiments, some areas of the simulation appear not to show any information registered; a zoom-in of an extreme case is illustrated in Fig. 4.

An example of the qualitative results of the accumulated images is shown in Fig. 5.

While this paper primarily focuses on the application of radiance fields for static scene reconstruction, it is important to note several limitations and potential avenues for future research.

Radiance fields have the ability to reconstruct dynamic scenes. The NeRF backbone utilized in our experiments did not have the capability to represent dynamic scenes, so we left their implementation as a task for future research.

Our simulator, by its design, does not rely on a specific representation of radiance fields. This flexibility allows for easy integration with alternative rendering techniques such as Gaussian splatting (Kerbl et al., 2023).

While our simulator demonstrates promising results in controlled environments, generalizing these findings to real-world applications presents additional challenges. Factors such as varying lighting conditions, occlusions, and reflective surfaces can significantly impact the performance and accuracy of radiance field reconstruction.

5 CONCLUSION

In this paper, we introduced a novel method for event camera simulation using neural radiance fields. Our approach leverages the capabilities of NeRFs to synthesize novel views of complex scenes, enabling the generation of realistic and diverse event camera data from arbitrary viewpoints. Experimental results demonstrate that our simulator matches or outperforms existing methods in terms of accuracy and realism, providing a valuable tool for the development and evaluation of event-based vision algorithms. The key contributions of this work include the integration of NeRFs with event-based sensing principles and the development of a versatile and efficient event camera simulator. We believe that this method represents a significant advancement in the field of event camera simulation, making this technology more accessible to researchers and developers.

REFERENCES

Brandli, C., Berner, R., Yang, M., Liu, S.-C., and Delbruck, T. (2014). A 240 × 180 130 db 3 µs latency global shutter spatiotemporal vision sensor. *IEEE Journal of Solid-State Circuits*, 49(10):2333–2341.

Cao, R., Galor, D., Kohli, A., Yates, J. L., and Waller, L. (2024). Noise2image: Noise-enabled static scene recovery for event cameras.

Chen, A., Xu, Z., Geiger, A., Yu, J., and Su, H. (2022).

Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision (ECCV)*.

García, G. P., Camilleri, P., Liu, Q., and Furber, S. (2016). pydvs: An extensible, real-time dynamic vision sensor emulator using off-the-shelf hardware. In *IEEE Symposium Series on Computational Intelligence*, pages 1–7.

Gehrig, D., Gehrig, M., Hidalgo-Carrió, J., and Scaramuzza, D. (2020). Video to events: Recycling video datasets for event cameras. In *IEEE/CVF Conference on Computer Vision and Pattcalern Recognition*.

Guo, S. and Delbruck, T. (2023). Low cost and latency event camera background activity denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):785–795.

Horé, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *In International Conference on Pattern Recognition*, pages 2366–2369.

Hu, Y., Liu, S. C., and Delbruck, T. (2021). v2e: From video frames to realistic DVS events. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE.

Hwang, I., Kim, J., and Kim, Y. M. (2023). Ev-nerf: Event based neural radiance field. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 837–847.

Joubert, D., Marcireau, A., Ralph, N., Jolley, A., van Schaik, A., and Cohen, G. (2021). Event camera simulator improvements via characterized parameters. *Frontiers in Neuroscience*, 15.

Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4).

Klenk, S., Koestler, L., Scaramuzza, D., and Cremers, D. (2023). E-nerf: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*.

Li, W., Saeedi, S., McCormac, J., Clark, R., Tzoumanikas, D., Ye, Q., Huang, Y., Tang, R., and Leutenegger, S. (2018). Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset. In *British Machine Vision Conference*.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020). Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*.

Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., and D. (2017). Scaramuzza: The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *International Journal of Robotics Research*, 36:142–149.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15.

Rebecq, H., Gehrig, D., and Scaramuzza, D. (2018). ESIM: an open event camera simulator. *Conference on Robot Learning (CoRL)*.

Rudnev, V., Elgharib, M., Theobalt, C., and Golyanik, V. (2023). Eventnerf: Neural radiance fields from a single colour event camera. In *Computer Vision and Pattern Recognition (CVPR)*.

Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Tagliasacchi, A. and Mildenhall, B. (2022). Volume rendering digest (for NeRF). *arXiv:2209. 02417 [cs]*, 3:02417.

Union, I. T. (2011). Recommendation itu-r bt.601-7: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. `https://www.itu.int/rec/R-REC-BT.601`.

Zhang, Z., Cui, S., Chai, K., Yu, H., Dasgupta, S., Mahbub, U., and Rahman, T. (2024). V2ce: Video to continuous events simulator.

Zhu, A. Z., Wang, Z., Khant, K., and Daniilidis, K. (2019). Eventgan: Leveraging large scale image datasets for event cameras. *arXiv preprint arXiv:1912.01584*.