HFR and HDR Video from Multi-Attenuated Spikes Using a Rapidly Rotating SpokeND Filter

Yakun Chang[3,4#] Zhaojun Huang[1,2#] Siqi Yang[1,2,5] Yeliduosi Xiaokaiti[1,2]
Shikui Wei[3,4] Yao Zhao[3,4] Tiejun Huang[1,2] Boxin Shi[1,2,6*]

[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University

[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[3] Institute of Information Science, Beijing Jiaotong University

[4] Visual Intelligence +X International Cooperation Joint Laboratory of the MoE

[5]Institute for Artificial Intelligence, Peking University, Peking University

[6] PKU-AI[2] Robotics Joint Lab of Embodied AI

{ykchang, shkwei, yzhao}@bjtu.edu.cn, {huangzhaojun, yongqiye}@stu.pku.edu.cn,
{yousiki, tjhuang, shiboxin}@pku.edu.cn

Abstract

Capturing scenes with both high dynamic range (HDR) and high-speed motion remains challenging for conventional cameras. Existing alternating-exposure approaches exacerbate temporal resolution loss, making them unsuitable for high-speed scenes. Consequently, current solutions typically fix spatial-varying attenuation levels or employ multiple sensors to maintain temporal resolution. In this paper, we leverage an ultra-high speed spike camera to enable spatial and temporal attenuation of incident light, thereby reconstructing high-frame-rate (HFR) and HDR video with a single sensor. We achieve this by placing a rapidly rotating spoke-pattern neutral density (SpokeND) filter in front of the camera, enabling each pixel to periodically capture multi-attenuated spikes. Building on these multi-attenuated spikes, we propose ReST-Net, which comprises the ReGain and ReFine modules. The ReGain module reconstructs spatially consistent frames by learning to recover relative gain from the multi-attenuated spikes, and the ReFine module removes temporal fluctuations to produce temporally consistent HDR videos. Extensive experiments on synthetic and real-world data demonstrate that our method can reconstruct HDR video at up to 2000 FPS.

1. Introduction

Real-world environments often present scenes with both high dynamic range (HDR) and high-speed motion. Conventional HDR video methods [3, 16, 18] which rely on frame-based cameras, face fundamental limitations in capturing high-speed scenes due to restricted frame rates and shutter speed constraints. In recent years, the neuromorphic spike cameras demonstrate inherent advantages through the high temporal resolution (20,000 Hz) and low redundancy (single-bit data) [14]. Unlike conventional frame-based cameras where all pixels share synchronized exposure, each pixel in a spike camera operates an independent photon accumulator. When the accumulated photoelectrons reach a specific threshold, the pixel triggers a spike signal of 1 and resets the accumulator. Since spikes are not directly interpretable by the human visual system, converting them into standard video representations is necessary for both human perception and downstream tasks [44, 47, 48].

The baseline image reconstruction with spikes exploits temporal accumulation over a time window [49, 50]. While longer time windows facilitate HDR performance [2], they introduce significant motion blur artifacts in high-speed scenes. To improve HDR performance under short time windows, modulating spike quantization levels [2] is effective, as higher quantization levels correspond to increased dynamic range. However, this method necessitates hardware modulation to the native spike sensor. In contrast to quantization-level modulation, the trigger threshold adjustment offers an alternative approach [51]. As shown in Fig. 1 (a), a small threshold results in dense spikes, enhancing sensitivity to incident photons, whereas a large threshold generates sparse spikes, helping to avoid saturation in bright regions. However, real-time threshold adjustment is infeasible in high-speed motion scenes.

Compared to internal sensor modifications, optical modulations of incident light offer a more flexible solution. Spatial modulations [29, 32, 33] utilize optical filters to achieve spatial-varying light attenuation. However, the fixed position of the filter in front of the sensor restricts each pixel

# Equal contribution. * Corresponding author.

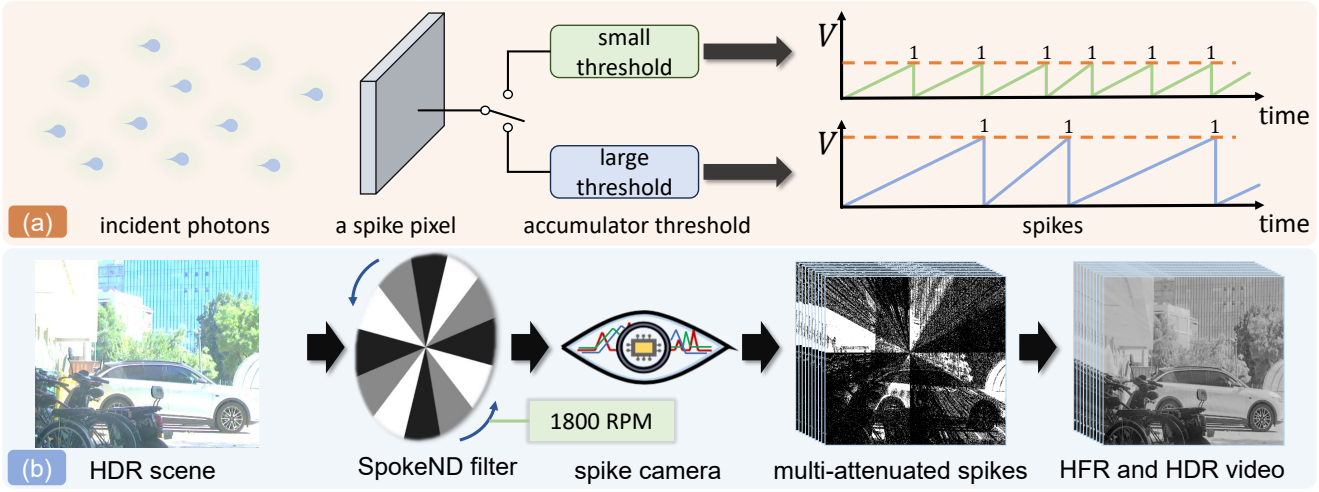

Figure 1. (a) When incident photons reach a pixel on a spike sensor, the photon-induced electrons generate a voltage V. Once V surpasses a threshold, the pixel triggers a spike 1. A small threshold (the orange dashed line) enhances sensitivity but increases the risk of saturated triggering. In contrast, a large threshold reduces the number of spikes, making it more suitable for bright scenes. (b) Since dynamically adjusting the threshold is impractical for high-speed scenes, we fix the spike camera to a small threshold and introduce a rapidly rotating (1800 revolutions per minute, RPM) spoke-pattern ND (SpokeND) Filter. Leveraging the ultra-high temporal resolution of the spike camera, our approach efficiently captures multi-attenuated spikes, enabling HDR video reconstruction in high-speed scenes.

to capturing light intensity at only one specific attenuation level. Consequently, reconstructing an HDR image requires spatial upsampling or interpolation to compensate spatial resolution. *If we can dynamically modulate the position of a spatial-varying filter, each pixel can capture multi-attenuated light intensities, thereby preserving spatial resolution.* However, in high-speed scenes, the low frame rates of conventional digital cameras pose a fundamental challenge in implementing such temporal modulation. For example, the LCD attenuator and electronics [31] can be used to control the incident light, this method achieves a frame rate of 30 FPS. While synchronized multi-sensor systems can capture multiple attenuation levels at 35 FPS [39], this approach remains insufficient for high-speed motion and introduces significant hardware complexity.

In this work, we exploit the ultra-high temporal resolution of a single spike camera to enable spatial and temporal modulation, which facilitates HDR video reconstruction up to 2000 FPS. As shown in Fig. 1 (b), a rapidly rotating spoke-pattern neutral density (**SpokeND**) filter is positioned in front of the spike camera, where the periodically distributed spokes exhibit multi-level light attenuation. This design allows each pixel to periodically receive multi-attenuated light during the rapid rotation. Meanwhile, the ultra-high temporal resolution ensures accurate sampling of these multi-attenuated spikes. However, this motivation faces challenges in high-speed scenes, where the incident light received by each pixel fluctuates over time, introducing difficulties in achieving temporal consistency. Moreover, recovering the gain, *i.e.*, removing the spatial-varying attenuation, is not a simple linear process, making the reconstruction of spatially consistent and perceptually pleasing videos non-trivial.

To address these issues, we propose a two-stage pipeline, named **ReST-Net**, which aims to **Re**construct high-frame-rate (HFR) and HDR videos by resolving the **S**patial variation and **T**emporal fluctuations in the multi-attenuated spikes. In the first stage, to handle the spatial variation, we introduce the **ReGain** module, which learns to **Re**cover the spatially consistent **Gain** of spikes relative to the non-attenuated condition, thereby reconstructing consistent frames. In the second stage, to suppress the temporal fluctuations between adjacent video frames, we design the **ReFine** module, which further refines the output to **Re**construct **Fine**r HDR videos with improved temporal consistency and visual coherence. To train and validate the proposed networks, we develop a customized simulator to synthesize multi-attenuated spikes. Furthermore, to demonstrate the effectiveness of our approach in real-world scenes, as shown in Fig. 3, we build the corresponding camera system equipped with our custom-designed SpokeND filter and collect 100 groups of real-world data. Our major contributions can be concluded as follows:

- A high-speed modulation scheme based on a rapidly rotating spoke pattern, enabling fast multi-attenuated sampling via the spike camera's high temporal resolution;
- A modulated spike camera system supported by a custom-designed SpokeND filter, allowing us to collect a sufficiently large dataset to support real-world data evaluation;
- A two-stage reconstruction pipeline, ReST-Net, to address spatial and temporal fluctuations and achieve robust HFR and HDR video reconstruction on real-world data.

2. Related work

HDR imaging with conventional sensors. Conventional sensors struggle to capture HDR ambient lighting within a single exposure. Reconstruction methods based on single low-dynamic-range (LDR) images [6, 8, 10, 20, 21, 23] often fail to recover fine details in severely underexposed or overexposed regions. To improve HDR performance in digital cameras, spatial modulations [29, 32, 33] use optical masks to achieve spatially varying light attenuation. However, the fixed mask position limits each pixel to a single attenuation level. In high-speed scenes, the low frame rates of conventional cameras hinder the implementation of temporal modulation. To control pixel exposures over time, Nayer *et al*. [31] use an LCD attenuator and electronics to control the incident light, this method only achieves a frame rate of 30 FPS. Although synchronized multi-sensor systems can capture multiple attenuation levels at 35 FPS [39], they remain inadequate for high-speed motion and introduce significant hardware complexity. An alternative approach involves fusing multiple LDR images taken with alternating exposures [7, 28], though this often introduces ghosting artifacts in dynamic scenes. To mitigate such artifacts and enhance image sharpness, image alignment techniques [24, 36] and deep learning methods [16, 41] have been proposed. Lee and Song [22] exploit motion information from high frame-rate sequences to improve HDR synthesis and suppress ghosting artifacts. Furthermore, merging alternating-exposure frame sequences has proven effective for reconstructing HDR videos at frame rates from 20 to 60 FPS [11, 17–19, 25, 26]. Chen *et al*. [3] enhance reconstruction quality by proposing a coarse-to-fine network that aligns and fuses in both image and feature spaces.

Video reconstruction with spikes. Reconstructing high-speed videos from spikes typically relies on the imaging model of spike cameras [9, 14, 45, 46]. Textures can be reconstructed either by accumulating spikes within a defined time window or by estimating the interval between successive spikes [49, 50]. However, due to the limited number of photons captured within ultra-short exposure times, such naive reconstruction suffers from considerable noise. To alleviate this issue, Chang *et al*. [2] propose improving the signal-to-noise ratio (SNR) by combining multi-bit and binary spikes through a rolling-mixed-bit scheme. However, this method requires hardware modifications, limiting its applicability in broader scenarios. Zhao *et al*. introduce Spk2ImgNet [43], a hierarchical network that progressively fuses spikes to enhance reconstruction quality. Despite its effectiveness, Spk2ImgNet exhibits limited generalization to real-world data due to its reliance on synthetic training datasets. To address the scarcity of real-world ground truth, self-supervised learning approaches [4, 5, 42] have been developed to reduce dependence on synthetic data. However, reconstructing HDR videos from spikes remains still challenging as the single-bit data is insufficient to cover the high dynamic range of real-world scenes. Chang *et al*. [1] propose a hybrid spike-RGB camera system that performs spatial alignment and frame interpolation simultaneously, enabling the recovery of 1000 FPS color videos. However, this hybrid system demands precise synchronization and optical alignment, and the additional space required by the beam splitter complicates the design of compact devices.

3. HFR and HDR approach

We propose a framework for reconstructing HFR and HDR video from multi-attenuated spikes. In Sec. 3.1, we introduce multi-attenuated spikes. In Sec. 3.2, we introduce our method for creating such spikes using a rapidly rotating SpokeND filter, along with the hardware platform design. In subsec:3.3, we present ReST-Net, a two-stage network that reconstruct HFR and HDR video.

3.1. Preliminaries on multi-attenuated Spikes

For each pixel in the spike camera, photo-generated electrons are continuously accumulated as long as the accumulated voltage V remains below the threshold V_{th}. Meanwhile, the readout circuit samples the pixel value at a fixed interval τ, outputting a value of 0 at each sampling point by default. Once V reaches or exceeds V_{th}, a spike signal of 1 is read out. We denote the accumulated voltage at a time t as $V(t)$, the corresponding spike signal $S(t)$ is

$$S(t) = \begin{cases} 1, & V(t) \geq V_{\text{th}}, \\ 0, & \text{otherwise}. \end{cases} \tag{1}$$

When measuring the dynamic range of a spike camera, it is necessary to accumulate spikes over a time window comprising N sampling intervals. Let I_{s} denote the reconstructed light intensity from spike accumulation, defined as:

$$I_{\text{s}}(t) = \eta_1 \sum_{i \in [0, N-1]} S(t + i\tau), \tag{2}$$

where η_1 is a proportionality constant. Similar to conventional sensors, the dynamic range of spike cameras is defined as $\text{DR} = 20 \log \frac{I_{\text{max}}}{I_{\text{min}}}$, where I_{max} and I_{min} represent the maximum and minimum detectable light intensities under the condition that the signal-to-noise ratio (SNR) is larger than 1 [2]. Without considering overexposure, it is evident that a larger N enables the spike camera to cover higher dynamic range, as both lower I_{min} and higher I_{max} can be detected. However, in high-speed scenes, a long accumulation window is not feasible, as it introduces significant motion blur. To address this, we configure the spike sensor for high sensitivity to the environment, *i.e*., small V_{th}, allowing the detection of a low I_{min}. However, the high sensitivity also raises the risk of overexposure in regions of strong light intensities, thereby limiting I_{max}. To

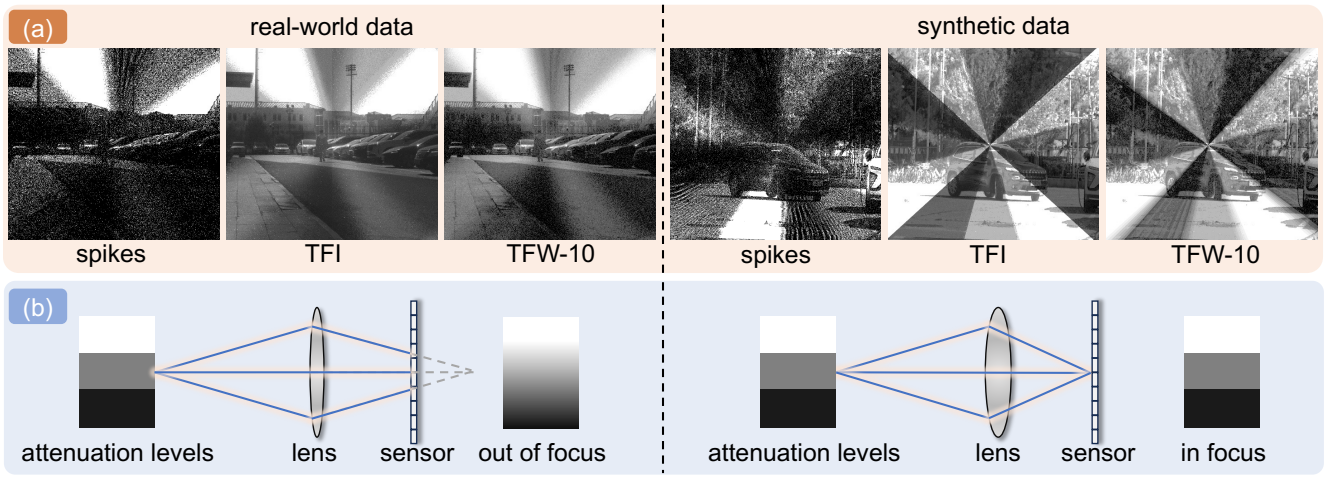

Figure 2. (a) Examples of our real-world and synthetic data. We show baseline image construction using TFI and TFW [14], where TFW-10 denotes a temporal window spanning 10 spike sampling intervals. (b) When the camera lens is focused on the scene, the modulation pattern introduced by the SpokeND filter may become slightly defocused. Despite the small gap between real-world and synthetic data, our method remains robust and is still able to reconstruct high-quality HFR and HDR videos without requiring retraining (See Fig. 6).

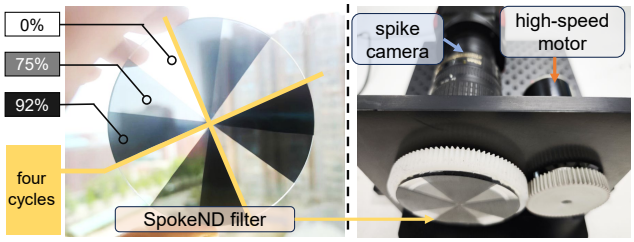

Figure 3. The custom-designed SpokeND filter and the dedicated hardware platform. SpokeND filter features three levels of attenuation, arranged across four repeating cycles. The filter is mounted on a gear and is rotated at a speed of 1800 RPM. Rotation is driven by a high-speed motor through the side-mounted ceramic bearing, which is lightweight and durable, ensuring stable and continuous operation. More details of the platform are available in the supplementary material.

extend the dynamic range while preserving sensitivity, we introduce multi-attenuated spikes. This can be achieved by rapidly and periodically modulating the attenuation of incident light: Higher attenuation levels prevent spike saturation in bright regions, while lower attenuation levels maintain sensitivity to I_{min}.

3.2. Creating multi-attenuated spikes

Multi-attenuated spikes can be collected by placing an electronically controlled LCD attenuator or a moving spatially varying filter in front of the spike camera. For simplicity and reliability, we design a rapidly rotating SpokeND filter, where each spoke corresponds to a distinct attenuation level. Meanwhile, since acquiring real-world ground truth is infeasible for high-speed scenes, we develop a new simulator to generate synthetic data. Samples from our dataset are shown in fig:defocus (a).

SpokeND filter. To achieve stable and high-speed multi-level attenuation, we carefully design a SpokeND filter tailored to our spike camera. As shown in Fig. 3, we address the following critical requirements for designing the SpokeND filter: (1) *Multiple attenuation levels.* Following the multi-exposure strategy commonly adopted in conventional HDR methods [16–18], the SpokeND filter incorporates three discrete attenuation levels, *i.e.*, 92%, 75%, and 0% [1], where the larger percentage indicates lower transmittance. (2) *High-frequency periodicity.* The attenuation regions are symmetrically distributed around the rotation center to preserve spatial balance. To further improve the temporal resolution of spike modulation and better adapt it to high-speed scenes, we design the filter with four repeated attenuation cycles. At a rotation speed of 1800 revolutions per minute (RPM), this configuration achieves a temporal modulation frequency of 7200 cycles per minute (CPM), significantly improving temporal modulation granularity. (3) *Lightweight and stability.* The filter is fabricated from optical resin, chosen for its lightweight nature and mechanical robustness.

Real-world data. As shown in Fig. 3, we successfully build the hardware platform capable of high-speed and multi-attenuated spike capture. The platform integrates a spike camera with our custom-designed SpokeND filter, all mounted on an optical breadboard. We developed a ceramic bearing with integrated gears to securely support the rotating SpokeND filter. A high-speed motor drives its rotation uniformly, reaching up to 1800 RPM, ensuring sta-

[1] Empirical setting. And for 0%, we neglect the intrinsic absorption of the filter material.

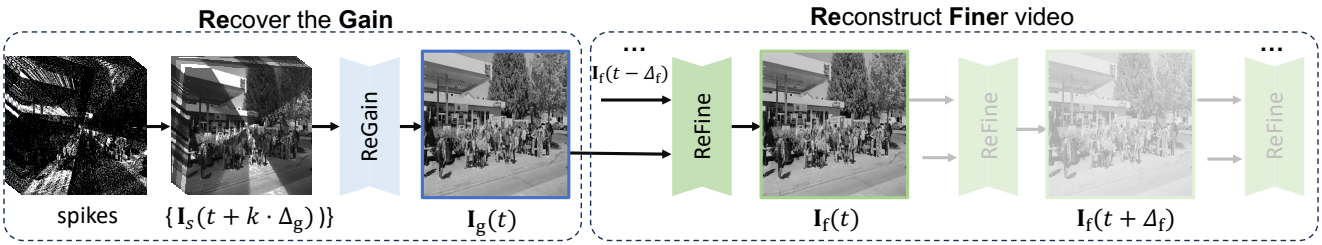

Figure 4. ReST-Net is a two-stage model that enables video reconstruction at arbitrary frame rates up to 2000 FPS. ReST-Net consists of two modules, *i.e.*, ReGain and ReFine, to reconstruct HFR and HDR video in a coarse-to-fine manner. ReGain module first performs a baseline reconstruction based on Eqns. (3), yielding the baseline frame denoted as $\mathbf{I}_s(t)$. To obtain the non-attenuated spike frame $\mathbf{I}_g(t)$, we collect a local temporal window of $2K+1$ baseline frames centered at t, with a fixed interval Δ_g between adjacent frames, and feed them into ReGain module. For temporal refinement, ReFine module reconstructs frame $\mathbf{I}_f(t)$ by incorporating consistency across time. Specifically, we concatenate the current ReGain output $\mathbf{I}_g(t)$ with a previous frame $\mathbf{I}_g(t-\Delta_f)$ and feed them into ReFine module. Here, Δ_f denotes the frame interval of the refined output.

ble and reliable high-speed operation. Using this platform, we collect 100 groups of real-world data for testing, including 80 captured in indoor scenes and the remaining in outdoor scenes. Each group of data contains one second of multi-attenuated spike recordings. Since capturing the corresponding non-attenuated spikes is not repeatable in real-world scenes, we only use this portion of the data for subjective evaluation.

Synthetic data. Since real-world data lacks corresponding ground truth, it cannot be used directly for model training. Therefore, we rely on synthetic data for training purposes. Each synthetic sample consists of three components: multi-attenuated spikes, non-attenuated spikes, and the corresponding ground truth HDR video. In this paper, we use the HDR videos provided by Chang *et al*. [1] and Su *et al*. [37] as our ground truth. During spike synthesis, we apply a uniform rotational speed to the simulated multi-attenuated SpokeND filter. The attenuation of light passing through the SpokeND filter is simulated by element-wise multiplication with attenuation levels. Based on the integrate-and-trigger mechanism of spike cameras [14], we set an accumulator and a trigger threshold, which allows us to generate the multi-attenuated and non-attenuated spikes. This dataset consists of 285 groups, with 235 groups for training and 50 groups for testing.

Discussion on lens defocus. In our current prototype, the filter is positioned in front of the lens. This external placement introduces a certain degree of defocusing, as shown in Fig. 2 (b), particularly when the filter is not formed exactly at the focal plane. Despite the actual attenuation levels deviating from the ideal predefined settings, *i.e.*, synthetic data, our method is still able to reconstruct high-quality HDR video on real-world data without the need for retraining. The results on real-world data are shown in Fig. 6. A more optimal configuration is placing the SpokeND filter behind the lens and in close proximity to the spike sensor. This setup requires more advanced hardware integration, which we plan to explore in future work.

3.3. Architecture of ReST-Net

Given a set of spikes, we begin with a baseline image reconstruction. As formulated in Eqns. (2), a common approach is to accumulate spikes over a temporal window, *i.e.*, TFW. However, this method is highly sensitive to the window size and often results in motion blur, especially in high-speed scenes. To mitigate this, we estimate pixel intensities based on the intervals between successive spikes, *i.e.*, TFI. Let $\delta(t)$ represent the time interval between two adjacent spikes, the corresponding pixel intensity is then obtained as

$$I_s(t) = \eta_2 / \delta(t), \tag{3}$$

where η_2 is a proportionality constant. We further extend this definition from the individual pixel to the full image and denote the reconstructed frame at time t as $\mathbf{I}_s(t)$. As illustrated in Fig. 4, the multi-attenuated frame $\mathbf{I}_s(t)$ exhibits noticeable spatial variations and temporal fluctuations due to the rapidly rotating SpokeND filter. To address these issues and reconstruct HFR and HDR videos, we propose a two-stage network, ReST-Net, composed of the ReGain and ReFine modules, which reconstruct the target video in a coarse-to-fine manner.

ReGain module. The ReGain module is designed to reconstruct spike frames as if they are captured without attenuation. This task is challenging since the relationship between multi-attenuated spikes and their non-attenuated counterparts is inherently nonlinear. As shown in Fig. 2 (a), under high photon flux, the sky regions with 0% attenuation may exhibit spike saturation, causing loss of linearity between photon count and spike density. Conversely, in low-light regions, areas with 92% attenuation may produce no spikes at all, making direct analytical recovery unreliable.

Fortunately, thanks to the rapid rotation of the SpokeND filter, even if saturation or zero-spike conditions occur at a specific time t, the temporal neighborhood modulated by multiple attenuation levels can still provide spikes captured under a suitable light intensity. Therefore, for

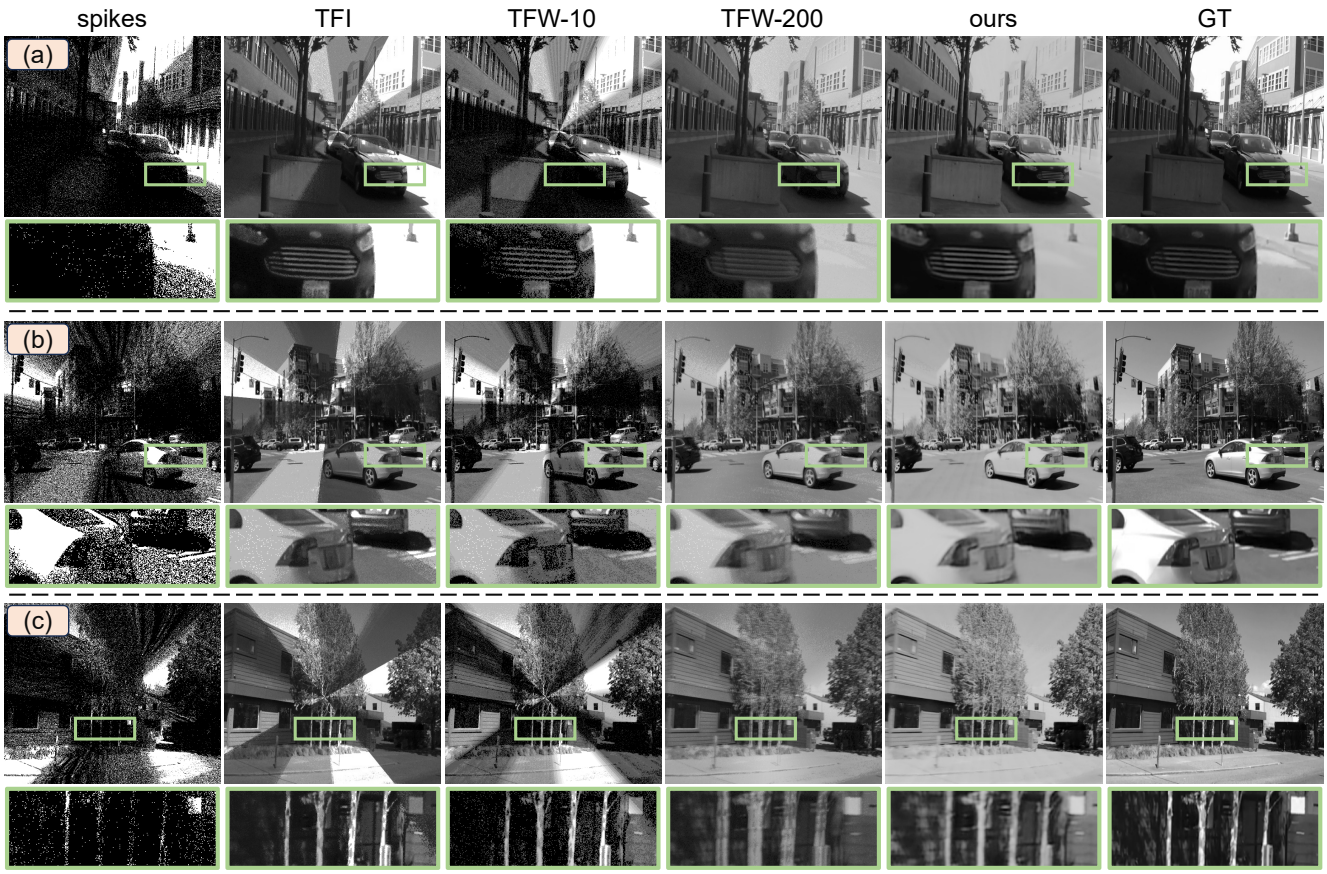

Figure 5. Visual equality comparison of synthetic data between the proposed method and compared methods. TFW-N indicates the TFW with a time window N. Please watch the videos in the supplementary material.

each inference step, the ReGain module takes a set of $2K+1$ temporally adjacent multi-attenuated spike frames as input, and the time interval between two adjacent frame is denoted by Δ_{g}. Thus, the input can be expressed as $\{\mathbf{I}_{\text{s}}(t+k\cdot\Delta_{\text{g}})\,|k\in[-K,K]\}$. Architecturally, ReGain adopts a U-Net [35] like encoder-decoder structure. To enhance its ability to capture spatial context, we integrate self-attention [38] blocks into the network, effectively expanding its receptive field. The output of our ReGain module is denoted as $\mathbf{I}_{\text{g}}(t)$.

ReFine module. ReFine module is designed to further refine the output of ReGain module and suppress inter-frame flickering. This module supports HFR video reconstruction at arbitrary frame rates up to 2000 FPS. Let Δ_{f} denote the target temporal interval between frames. To refine and output the frame $\mathbf{I}_{\text{f}}(t)$, we concatenate the current ReGain output $\mathbf{I}_{\text{g}}(t)$ with a previous frame $\mathbf{I}_{\text{g}}(t-\Delta_{\text{f}})$, and feed them into the ReFine module. This configuration provides consistency-aware guidance during refinement, helping to enhance temporal stability and reduce perceptual flicker in the final HDR video output. Similar to ReGain module, ReFine module adopts a U-Net like architecture enhanced with self-attention blocks.

Loss and training. The objective of the ReGain module is to recover spatially consistent intensity frames from the input multi-attenuated frame. To train this module, we first obtain the ground truth by applying Eqns. (3) to perform baseline reconstruction using synthetic non-attenuated spikes. The resulting ground truth for $\mathbf{I}_{\text{g}}(t)$ is denoted as $\mathbf{G}_{\text{g}}(t)$. The loss function used for training is defined as:$L_{\text{g}}(\mathbf{I}_{\text{g}},\mathbf{G}_{\text{g}})=\alpha_1 L_1+\alpha_2 L_2$, where L_1 and L_2 denote the pixel-wise ℓ_1 and ℓ_2 losses, respectively, and α_1, α_2 are weighting coefficients to balance the two terms. For the ReFine module, the ground truth is the HDR video frame used to synthesize the spike data, denoted as $\mathbf{G}_{\text{f}}(t)$. The loss function is defined as: $L_{\text{f}}=\beta_1 L_1+\beta_2 L_2+\beta_3 L_{\text{temp}}$, where β_1, β_2, and β_3 are weighting coefficients, L_{temp} is the temporal consistency loss, defined as: $L_{\text{temp}}=\ell_2(\mathbf{I}_{\text{f}}(t)-\mathbf{I}_{\text{f}}(t-\Delta_{\text{f}}),\mathbf{G}_{\text{f}}(t)-\mathbf{G}_{\text{f}}(t-\Delta_{\text{f}}))$. More details about the training are available in the supplementary material.

4. Experiments

We conduct experiments and evaluations on both synthetic and real-world datasets. To evaluate the effectiveness of our

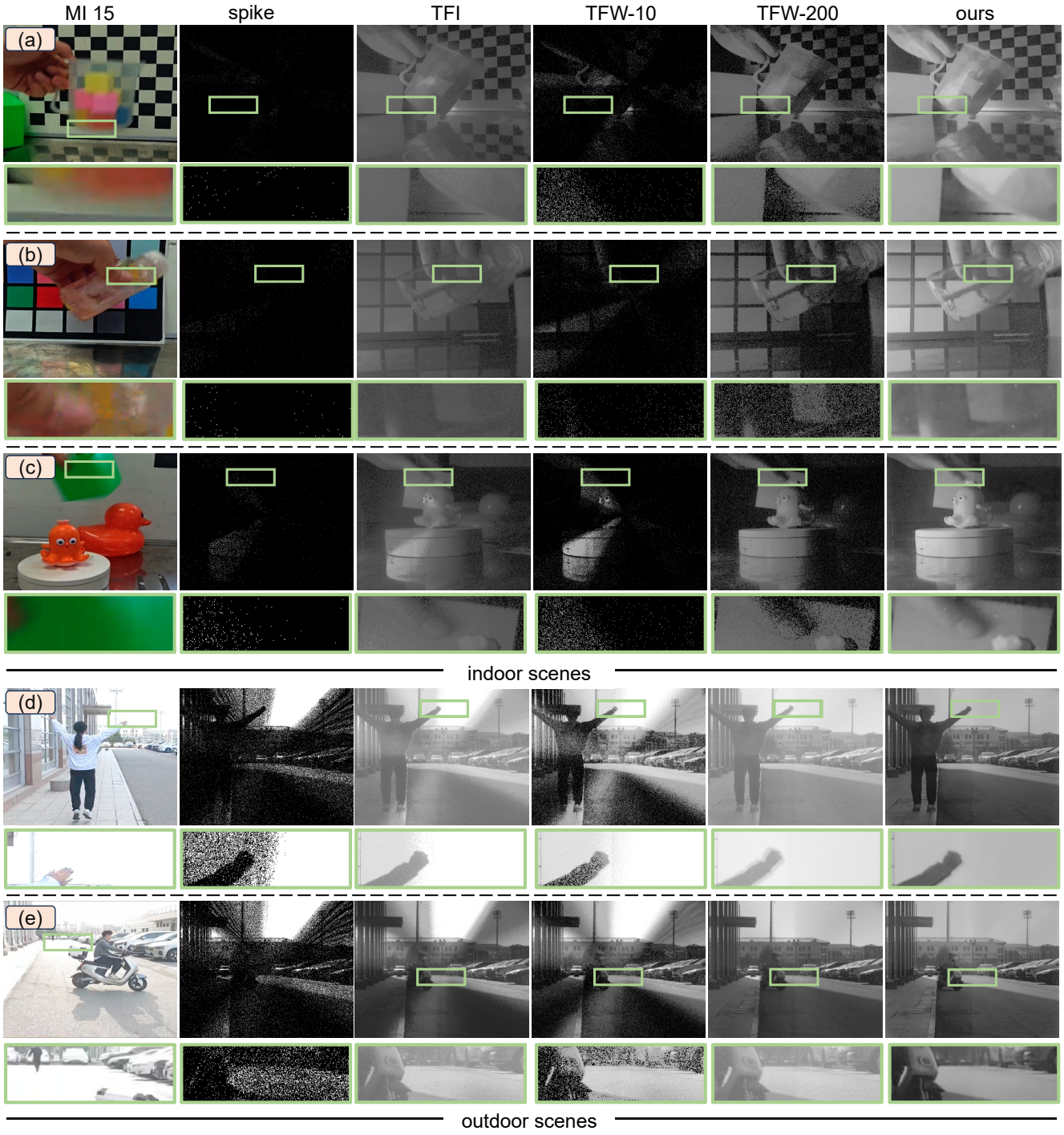

Figure 6. Visual equality comparison of real-world data between the proposed method and compared methods. For both indoor and outdoor scenes, our method demonstrates superior effectiveness in HDR video reconstruction. Please watch the videos in the supplementary material.

method, we compare it with existing video reconstruction methods based on spike camera, *i.e.*, TFW and TFI [14]. To the best of our knowledge, the proposed method is the first framework to reconstruct HFR and HDR videos with multi-attenuated spikes. While the compared methods are specifically designed for non-attenuated spikes, making the comparison not entirely fair, they still provide strong baselines to demonstrate the benefits of introducing approach.

Table 1. Quantitative results and ablation studies on our synthetic data. ↑ (↓) indicates larger (smaller) values are better.

	Comparison with baseline methods					Ablation study	
Method	TFW-10	TFW-70	TFW-200	TFI	Ours	w/o ReGain	w/o ReFine
PSNR↑	13.47	22.16	29.05	21.75	**34.27**	21.89	31.48
SSIM ↑	0.217	0.541	0.742	0.645	**0.916**	0.789	0.900
HDR-VDP3↑	2.858	4.508	6.313	4.804	**7.501**	4.859	7.173
HDR-VQM↓	1.530	0.797	0.289	0.736	**0.152**	0.724	0.166

4.1. Evaluation on synthetic data

We conducted both qualitative and quantitative evaluations on synthetic data. As shown in Fig. 5 (a), the sky and road exhibit high brightness, leading to spike saturation in non-attenuated regions. Our method successfully reconstructs fine details in these regions. For the moving car, our method generates blur-free frames. In contrast, both TFI and TFW-10 fail to recover spatially consistent frames. Although TFW-200 achieves spatial consistency through temporal averaging, it introduces noticeable motion blur in the moving car. In Fig. 5 (b), our method also effectively recovers details in saturated regions such as the sky and the white moving car. In Fig. 5 (c), our method preserves fine details of the trees, whereas TFI and TFW-10 exhibit significant noise and spatial inconsistency. Similarly, TFW-200 introduces pronounced blurring artifacts. We further evaluate the reconstructed HDR videos in terms of PSNR, SSIM [40], HDR-VDP3 [27] and HDR-VQM [30] in Table 1, showing that our method consistently achieves best performance.

Ablation study. To validate the effectiveness of the two modules, we conducted ablation studies on synthetic data *i.e.*, "w/o ReGain" and "w/o ReFine". As shown in Table 1, the quantitative results confirm that both modules contribute significantly to overall performance improvement. Fig. 7 presents two visual examples: Removing the ReFine module leads to diminished noise suppression, while removing the ReGain module results in a failure to reconstruct spatially consistent frames.

4.2. Evaluation on real-world data

We conduct qualitative comparisons on real-world data to demonstrate the effectiveness. As ground truth is unavailable in such settings, we perform qualitative evaluation only. As shown in Fig. 6, evaluations are carried out on both indoor and outdoor scenes. A commercial smartphone, *i.e.*, MI 15, is used to provide reference images of the HDR scenes. Note that the reference images are for illustrative purposes only and are not aligned with the spike camera. In Fig. 6 (a), (b), and (c), we introduce rapid object motions in indoor scenes. While existing methods suffer from significant noise and motion blur, our approach effectively reconstructs fine texture details. For outdoor scenes shown in Fig. 6 (d) and (e), reference images captured by the smartphone exhibit overexposure in the sky regions. Similarly, in the spike data, the 0% attenuation level leads to saturated triggering in those bright areas. Our method successfully reconstructs both the bright sky and the jumping person. In contrast, reconstructions from TFI and TFW-10 suffer from spatial inconsistencies and noise, while TFW-200 yields blurred results.

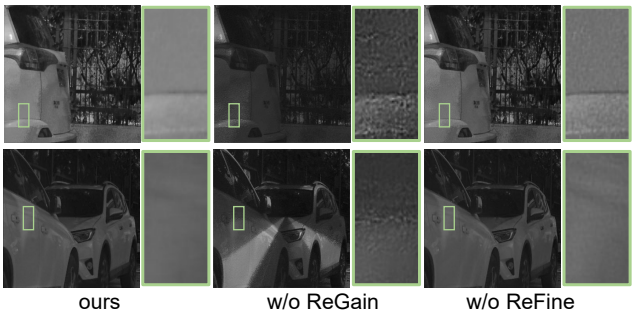

Figure 7. Visual comparison from the ablation study. Removing the ReGain module leads to persistent spatial variations that are difficult to correct, while removing the ReFine module results in reduced noise suppression.

5. Conclusion

In this paper, we present a novel framework for reconstructing HFR and HDR video from spikes modulated by a rapidly rotating SpokeND filter. By exploiting the ultra-high temporal resolution of spike cameras, our method introduces a spatiotemporal modulation mechanism that enables effective sampling of multi-attenuated spikes. To support this, we develop a dedicated hardware platform comprising a spike camera and a custom-designed SpokeND filter with four repeated attenuation cycles. Mounted on a high-speed rotating stage, the filter enables high-frequency, periodic modulation of incoming light, effectively encoding HDR information over time. Building upon this setup, we propose a two-stage ReST-Net, consisting of ReGain and ReFine modules, to progressively reconstruct HDR video in a coarse-to-fine manner. Extensive experiments demonstrate that our approach outperforms conventional spike-based reconstruction methods in HDR scenes.

Limitation and future work. Despite the promising results demonstrated by our approach, the limitation on motion speed still remains. To further clarify the robustness under varying motion speeds, we conduct additional experiments on the synthetic dataset. When motion speed is doubled, the results achieve PSNR = 29.91 and SSIM = 0.902. When tripled, PSNR = 26.35 and SSIM = 0.805. This demonstrates a predictable decline in quantitative scores with increasing motion speed, reflecting the theoretical limits.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Grant No. 62301009), Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012), and Beijing Natural Science Foundation (Grant No. L233024).

References

[1] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 1000 FPS HDR video with a spike-rgb hybrid camera. In *Proc. of Computer Vision and Pattern Recognition*, pages 22180–22190, 2023. 3, 5, 1

[2] Yakun Chang, Yeliduosi Xiaokaiti, Yujia Liu, Bin Fan, Zhaojun Huang, Tiejun Huang, and Boxin Shi. Towards HDR and HFR video from rolling-mixed-bit spikings. In *Proc. of Computer Vision and Pattern Recognition*, pages 25117–25127, 2024. 1, 3

[3] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K Wong, and Lei Zhang. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proc. of International Conference on Computer Vision*, pages 2502–2511, 2021. 1, 3

[4] Shiyan Chen, Chaoteng Duan, Zhaofei Yu, Ruiqin Xiong, and Tiejun Huang. Self-supervised mutual learning for dynamic scene reconstruction of spiking camera. page 2859–2866, 2022. 3

[5] Shiyan Chen, Zhaofei Yu, and Tiejun Huang. Self-supervised joint dynamic scene reconstruction and optical flow estimation for spiking camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 350–358, 2023. 3

[6] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. HDRUNet: Single image HDR reconstruction with denoising and dequantization. In *Proc. of Computer Vision and Pattern Recognition*, pages 354–363, 2021. 3

[7] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of ACM SIGGRAPH*, pages 1–10, 2008. 3

[8] Sebastian Dille, Chris Careaga, and Yağız Aksoy. Intrinsic single-image HDR reconstruction. In *Proc. of European Conference on Computer Vision*, pages 161–177. Springer, 2024. 3

[9] Yanchen Dong, Ruiqin Xiong, Jing Zhao, Jian Zhang, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. Joint demosaicing and denoising for spike camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 1582–1590, 2024. 3

[10] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics*, 36(6):1–15, 2017. 3

[11] Yulia Gryaditskaya, Tania Pouli, Erik Reinhard, Karol Myszkowski, and Hans-Peter Seidel. Motion aware exposure bracketing for HDR video. In *Proc. of Computer Graphics Forum*, pages 119–130, 2015. 3

[12] Liwen Hu, Rui Zhao, Ziluo Ding, Lei Ma, Boxin Shi, Ruiqin Xiong, and Tiejun Huang. Optical flow estimation for spiking camera. In *Proc. of Computer Vision and Pattern Recognition*, 2022. 1

[13] Liwen Hu, Lei Ma, Yijia Guo, and Tiejun Huang. SCSim: A realistic spike cameras simulator. In *Proc. of International Conference on Multimedia and Expo*, 2024. 1

[14] Tiejun Huang, Yajing Zheng, Zhaofei Yu, Rui Chen, Yuan Li, Ruiqin Xiong, Lei Ma, Junwei Zhao, Siwei Dong, Lin Zhu, et al. 1000× faster camera and machine vision with ordinary devices. *Engineering*, 2022. 1, 3, 4, 5, 7, 2

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning*, 2015. 1

[16] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144–1, 2017. 1, 3, 4

[17] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep HDR video from sequences with alternating exposures. In *Proc. of Computer graphics forum*, pages 193–205, 2019. 3

[18] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen. Patch-based high dynamic range video. *ACM Transactions on Graphics*, 32(6):202–1, 2013. 1, 4

[19] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, 2003. 3

[20] Joonsoo Kim, Zhe Zhu, Tien Bau, and Chenguang Liu. DCDR-UNet: Deformable convolution based detail restoration via u-shape network for single image HDR reconstruction. In *Proc. of Computer Vision and Pattern Recognition*, pages 5909–5918, 2024. 3

[21] Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua. Single-image HDR reconstruction by multi-exposure generation. In *Proc. of Winter Conference on Applications of Computer Vision*, pages 4063–4072, 2023. 3

[22] Byungju Lee and Byung Cheol Song. Multi-image high dynamic range algorithm using a hybrid camera. *Signal Processing: Image Communication*, 30:37–56, 2015. 3

[23] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proc. of Computer Vision and Pattern Recognition*, pages 1651–1660, 2020. 3

[24] Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Transactions on Image Processing*, 26(5):2519–2532, 2017. 3

[25] Stephen Mangiat and Jerry Gibson. High dynamic range video with ghost removal. In *Proc. of Applications of Digital Image Processing*, pages 307–314. SPIE, 2010. 3

[26] Stephen Mangiat and Jerry Gibson. Spatially adaptive filtering for registration artifact removal in HDR video. In *Proc. of International Conference on Image Processing*, pages 1317–1320, 2011. 3

[27] Rafal K Mantiuk, Dounia Hammou, and Param Hanji. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625*, 2023. 8

[28] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Proc. of Pacific Conference on Computer Graphics and Applications*, pages 382–390, 2007. 3

[29] Srinivasa G Narasimhan and Shree K Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005. 1, 3

[30] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet. HDR-VQM: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46–60, 2015. 8

[31] Nayar and Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1168–1175. IEEE, 2003. 2, 3

[32] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. of Computer Vision and Pattern Recognition*, pages 472–479, 2000. 1, 3

[33] Yutaro Okamoto, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi. Deep snapshot HDR imaging using multi-exposure color filter array. *The Visual Computer*, 40 (5):3285–3301, 2024. 1, 3

[34] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Proc. of Advances in Neural Information Processing Systems*. 2019. 1

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6, 1

[36] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):203–1, 2012. 3

[37] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proc. of Computer Vision and Pattern Recognition*, 2017. 5

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Proc. of Advances in Neural Information Processing Systems*, 30, 2017. 6, 1

[39] Hongcheng Wang, Ramesh Raskar, and Narendra Ahuja. High dynamic range video using split aperture camera. In *IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Washington, DC, USA*. Citeseer, 2005. 2, 3

[40] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 8

[41] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep HDR imaging via a non-local network. *IEEE Transactions on Image Processing*, 29: 4308–4322, 2020. 3

[42] Siqi Yang, Zhaojun Huang, Yakun Chang, Bin Fan, Zhaofei Yu, and Boxin Shi. Real-data-driven 2000 FPS color video from mosaicked chromatic spikes. In *Proc. of European Conference on Computer Vision*, pages 1111–2222, 2024. 3

[43] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proc. of Computer Vision and Pattern Recognition*, pages 11996–12005, 2021. 3

[44] Jing Zhao, Ruiqin Xiong, Jiyu Xie, Boxin Shi, Zhaofei Yu, Wen Gao, and Tiejun Huang. Reconstructing clear image for high-speed motion scene with a retina-inspired spike camera. *IEEE Transactions on Computational Imaging*, 8:12–27, 2021. 1

[45] Jing Zhao, Ruiqin Xiong, Jian Zhang, Rui Zhao, Hangfan Liu, and Tiejun Huang. Learning to super-resolve dynamic scenes for neuromorphic spike camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 3579–3587, 2023. 3

[46] Junwei Zhao, Jianming Ye, Shiliang Shiliang, Zhaofei Yu, and Tiejun Huang. Recognizing high-speed moving objects with spike camera. In *Proc. of ACM MM*, pages 7657–7665, 2023. 3

[47] Rui Zhao, Ruiqin Xiong, Jian Zhang, Zhaofei Yu, Shuyuan Zhu, Lei Ma, and Tiejun Huang. Spike camera image reconstruction using deep spiking neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(6): 5207–5212, 2023. 1

[48] Yajing Zheng, Lingxiao Zheng, Zhaofei Yu, Boxin Shi, Yonghong Tian, and Tiejun Huang. High-speed image reconstruction through short-term plasticity for spiking cameras. In *Proc. of Computer Vision and Pattern Recognition*, pages 6358–6367, 2021. 1

[49] Lin Zhu, Siwei Dong, Tiejun Huang, and Yonghong Tian. A retina-inspired sampling method for visual texture reconstruction. In *Proc. of International Conference on Multimedia and Expo*, pages 1432–1437, 2019. 1, 3

[50] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proc. of Computer Vision and Pattern Recognition*, pages 1438–1446, 2020. 1, 3

[51] Zhenkun Zhu, Ruiqin Xiong, Jing Zhao, Rui Zhao, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. High dynamic range imaging for dynamic scenes based on multi-level spike camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 1

HFR and HDR Video from Multi-Attenuated Spikes Using a Rapidly Rotating SpokeND Filter

Supplementary Material

Yakun Chang[3,4#] Zhaojun Huang[1,2#] Siqi Yang[1,2,5] Yeliduosi Xiaokaiti[1,2]
Shikui Wei[3,4] Yao Zhao[3,4] Tiejun Huang[1,2] Boxin Shi[1,2,6*]
[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[3] Institute of Information Science, Beijing Jiaotong University
[4] Visual Intelligence +X International Cooperation Joint Laboratory of the MoE
[5]Institute for Artificial Intelligence, Peking University, Peking University
[6] PKU-AI[2] Robotics Joint Lab of Embodied AI
{ykchang, shkwei, yzhao}@bjtu.edu.cn, {huangzhaojun, yongqiye}@stu.pku.edu.cn,
{yousiki, tjhuang, shiboxin}@pku.edu.cn

In the supplementary material, we provide details of our hardware platform in Sec. 5.1, details of our simulator in Sec. 5.2, additional implementation details in Sec. 5.3, and additional qualitative results in Sec. 5.4.

5.1. Details of hardware platform

In this approach, we utilize the "Spike M1K40-H2-Gen3" spike camera, which supports a sampling rate of 20,000 Hz. The high-speed motor, "SOYG 3420 24V", drives the filter rotation at 1800 RPM. The SpokeND filter is custom-fabricated with a 10 mm diameter, constructed by attaching multiple optical films with varying attenuation levels onto a transparent optical resin.

5.2. Details of simulator

Existing spike simulators [1, 2, 12, 13] are not designed to support multi-attenuated spike generation. To accommodate our experimental setup, we develop the multi-attenuated spike simulator tailored to our spoke-pattern attenuation model. This simulator takes video datasets as input and applies tone mapping to simulate realistic illumination in HDR scenes. To introduce multiple attenuation levels, we firstly generate the spoke-pattern mask according to our custom-designed SpokeND filter. The mask is then rotated by an angular step consistent with the actual rotational speed of the SpokeND filter to reflect temporal variation. For the incident light intensity, *i.e.*, the pixel value of the ground truth frame, is modulated by the mask at each pixel. For the spike generation, we set the sampling interval parameter to determine how many spike frames each video frame represents.

To bridge the domain gap between synthetic and real-world data, we augment the synthetic data by applying random Gaussian blur to each spoke-pattern mask, simulating the lens defocus discussed in Section 3.2. Additionally, similar to the prior work [13], we incorporate both temporal and spatial noise simulations to enhance the realism of the synthetic spikes.

5.3. Additional implementation details

The ReGain module employs a modified U-Net [35] backbone enhanced with the self-attention mechanism. The encoder comprises a series of convolutional blocks with progressively increasing channel depth ($32 \rightarrow 64 \rightarrow 128 \rightarrow 256$), where downsampling is performed via strided convolutions. Each encoder layer consists of two convolutional blocks with Leaky ReLU activation. The bottleneck integrates a multi-head self-attention [38] mechanism to expand the receptive field, followed by two additional convolutional layers for deep feature refinement. In the decoder, skip connections are employed to facilitate high-fidelity spatial reconstruction. The ReFine module shares similar structures with ReGain module, The encoder consists of four convolutional stages with increasing channel depth ($4 \rightarrow 8 \rightarrow 16 \rightarrow 32$). Each encoder block stacks two convolutional layers with leaky ReLU activation and batch normalization [15]. The decoder mirrors the encoder in structure and includes bilinear upsampling followed by convolution, with skip connections from the encoder. The final output is produced via a 1×1 convolution followed by an upsampling layer to match the target resolution. Regarding the parameter settings in this approach, Δ_g is set to 0.5 ms, and K is empirically chosen as 4. Δ_f can be configured to any value greater than 0.5 ms, allowing the frame rate to reach up to 2000 FPS.

ReST-Net is implemented using PyTorch [34] and trained on a single NVIDIA RTX 4090 GPU. We first train the ReGain module for 50 epochs, followed by training the ReFine module for 30 epochs. During both training stages, we use a batch size of 8 and adopt the Adam optimizer with

# Equal contribution. * Corresponding author.

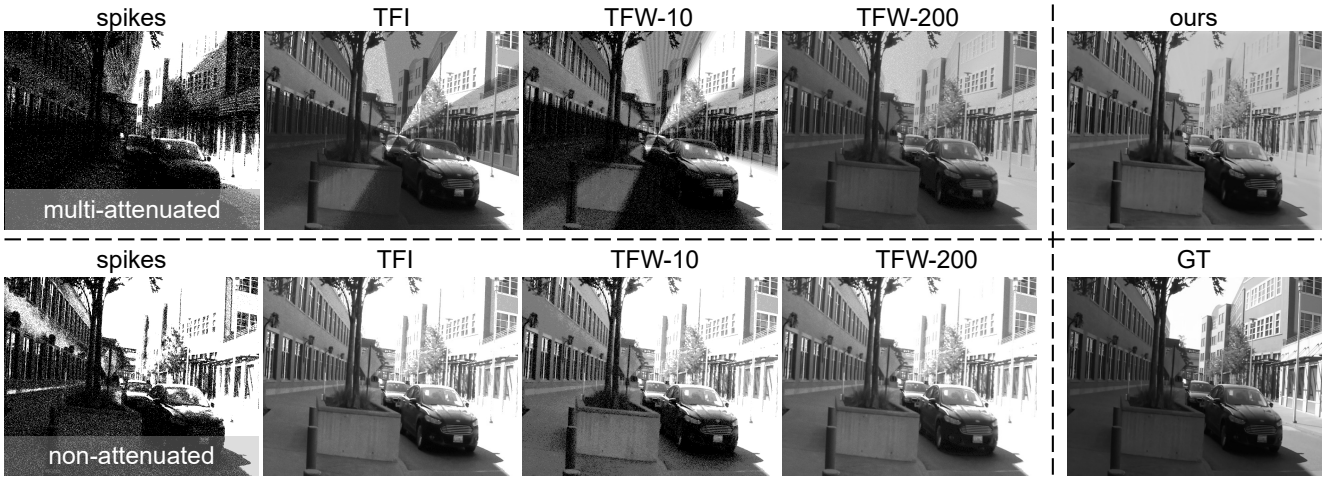

Figure 8. For Fig. 5 (a) in the main paper, we additionally simulate multi-attenuated and non-attenuated spikes to show the effectiveness of the SpokeND filter. The images reconstructed form non-attenuated spikes suffer from over-saturation.

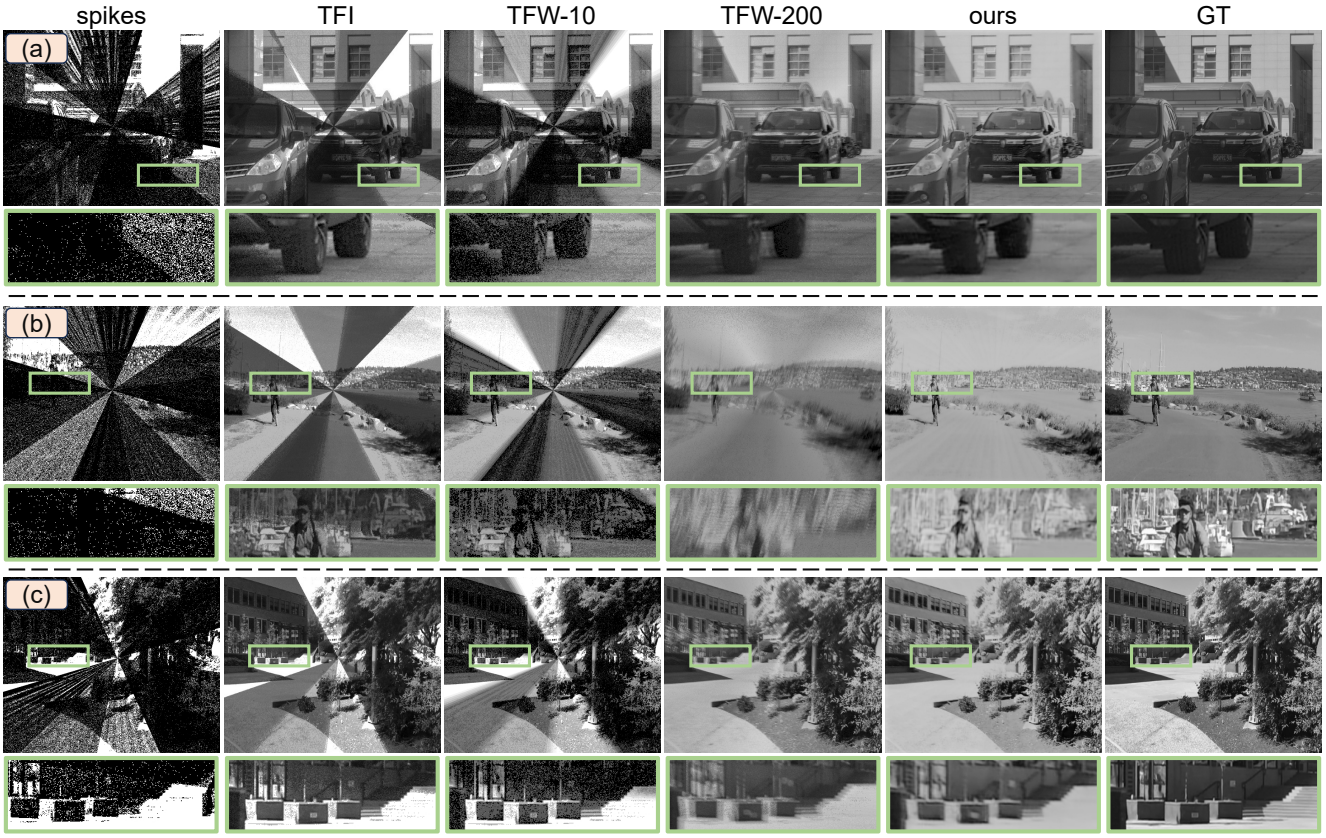

Figure 9. Additional visual equality comparison of synthetic data between the proposed method and compared methods, *i.e.*, TFI and TFW [14].The ground truth image is tone-mapped for better visualization.

an initial learning rate of 0.0001. α_1 and α_2 are set to be 0.4 and 1.0. β_1, β_2, and β_3 are set to 0.4, 1.0 and 1.0, respectively.

The average inference time of ReST-Net is 21.53 ms per frame, corresponding to an overall output rate of 46.44 FPS. In terms of model size, the ReGain module contains 19.3 million parameters (73.74 MB), while the ReFine module comprises 449K parameters (1.72 MB), demonstrating the

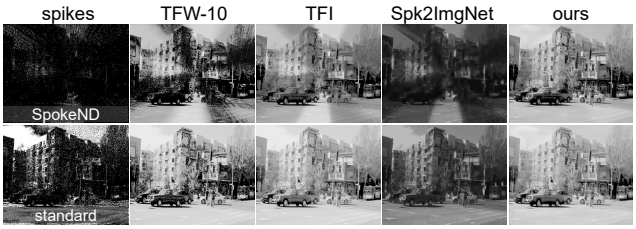

Figure 10. Additional visual equality comparison of synthetic data between the proposed method and compared methods, *i.e.*, TFI, TFW [14], and Spk2ImgNet [43].

efficiency of our two-stage architecture.

5.4. Additional qualitative results

Ablation study. To further validate the necessity of the SpokeND filter, we simulate HDR scenes as shown in Fig. 8. Without the SpokeND filter, the non-attenuated spikes exhibit saturated triggering in high-intensity regions. Under such conditions, the loss of information makes it impossible for all the methods to reconstruct textures in the saturated areas. More detailed ablation studies in Table 2 demonstrate effectiveness of attention module and the select of time window. Removing the attention module leads to noticeable performance degradation. Moreover, the input layer is designed to process a time window of $2K+1$ spike frames, spanning 10 ms to cover a full attenuation cycle.

Table 2. More ablation studies.

Table A. Detailed ablation study

Metrics	Ours($K=9$)	w/o ReGain	w/o ReFine	w/o Attention	$K=1$	$K=3$
PSNR↑	34.27	21.89	31.48	32.34	28.04	30.81
SSIM ↑	0.916	0.789	0.900	0.914	0.872	0.873

Compare with existing methods. We show more comparison results with existing methods on synthetic data in Fig. 9. As shown, our method reconstructs sharp video frames. Although TFW-200 [14] can also mitigate the spatial variation introduced by the rotating filter through temporal averaging, it tends to produce noticeably blurred results. We further demonstrate the utility of the SpokeND filter by employing standardized spike data for fair comparison with one state-of-the-art reconstruction method (*e.g.*, Spk2ImgNet [43]). As shown in Fig. 10, Spk2ImgNet [43] introduces motion-blur artifacts and noticeable noise, whereas our method produces cleaner reconstructions.

Motion limits. Aliasing and harmonic synchronization pose challenges for mechanical modulation. As illustrated in Fig. 11, we conduct experiments with a fan at 3 rotation speeds (240, 600, and 1800 RPM) to further investigate the motion limits: Our method is capable of handling a fan rotation speed of 240 RPM, where the fan rotates 1° in just 0.7 ms.

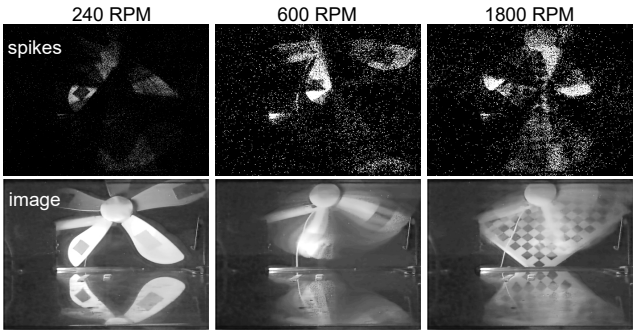

Figure 11. Test on a high-speed fan: Reconstructed results show motion blur at ultra-high rotational speeds.

References

[1] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 1000 FPS HDR video with a spike-rgb hybrid camera. In *Proc. of Computer Vision and Pattern Recognition*, pages 22180–22190, 2023. 3, 5, 1

[2] Yakun Chang, Yeliduosi Xiaokaiti, Yujia Liu, Bin Fan, Zhaojun Huang, Tiejun Huang, and Boxin Shi. Towards HDR and HFR video from rolling-mixed-bit spikings. In *Proc. of Computer Vision and Pattern Recognition*, pages 25117–25127, 2024. 1, 3

[3] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K Wong, and Lei Zhang. HDR video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proc. of International Conference on Computer Vision*, pages 2502–2511, 2021. 1, 3

[4] Shiyan Chen, Chaoteng Duan, Zhaofei Yu, Ruiqin Xiong, and Tiejun Huang. Self-supervised mutual learning for dynamic scene reconstruction of spiking camera. page 2859–2866, 2022. 3

[5] Shiyan Chen, Zhaofei Yu, and Tiejun Huang. Self-supervised joint dynamic scene reconstruction and optical flow estimation for spiking camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 350–358, 2023. 3

[6] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. HDRUNet: Single image HDR reconstruction with denoising and dequantization. In *Proc. of Computer Vision and Pattern Recognition*, pages 354–363, 2021. 3

[7] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of ACM SIGGRAPH*, pages 1–10, 2008. 3

[8] Sebastian Dille, Chris Careaga, and Yağız Aksoy. Intrinsic single-image HDR reconstruction. In *Proc. of European Conference on Computer Vision*, pages 161–177. Springer, 2024. 3

[9] Yanchen Dong, Ruiqin Xiong, Jing Zhao, Jian Zhang, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. Joint demosaicing and denoising for spike camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 1582–1590, 2024. 3

[10] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K

Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM Transactions on Graphics*, 36(6):1–15, 2017. 3

[11] Yulia Gryaditskaya, Tania Pouli, Erik Reinhard, Karol Myszkowski, and Hans-Peter Seidel. Motion aware exposure bracketing for HDR video. In *Proc. of Computer Graphics Forum*, pages 119–130, 2015. 3

[12] Liwen Hu, Rui Zhao, Ziluo Ding, Lei Ma, Boxin Shi, Ruiqin Xiong, and Tiejun Huang. Optical flow estimation for spiking camera. In *Proc. of Computer Vision and Pattern Recognition*, 2022. 1

[13] Liwen Hu, Lei Ma, Yijia Guo, and Tiejun Huang. SCSim: A realistic spike cameras simulator. In *Proc. of International Conference on Multimedia and Expo*, 2024. 1

[14] Tiejun Huang, Yajing Zheng, Zhaofei Yu, Rui Chen, Yuan Li, Ruiqin Xiong, Lei Ma, Junwei Zhao, Siwei Dong, Lin Zhu, et al. 1000× faster camera and machine vision with ordinary devices. *Engineering*, 2022. 1, 3, 4, 5, 7, 2

[15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning*, 2015. 1

[16] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144–1, 2017. 1, 3, 4

[17] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep HDR video from sequences with alternating exposures. In *Proc. of Computer graphics forum*, pages 193–205, 2019. 3

[18] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B Goldman, and Pradeep Sen. Patch-based high dynamic range video. *ACM Transactions on Graphics*, 32(6):202–1, 2013. 1, 4

[19] Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High dynamic range video. *ACM Transactions on Graphics*, 22(3):319–325, 2003. 3

[20] Joonsoo Kim, Zhe Zhu, Tien Bau, and Chenguang Liu. DCDR-UNet: Deformable convolution based detail restoration via u-shape network for single image HDR reconstruction. In *Proc. of Computer Vision and Pattern Recognition*, pages 5909–5918, 2024. 3

[21] Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua. Single-image HDR reconstruction by multi-exposure generation. In *Proc. of Winter Conference on Applications of Computer Vision*, pages 4063–4072, 2023. 3

[22] Byungju Lee and Byung Cheol Song. Multi-image high dynamic range algorithm using a hybrid camera. *Signal Processing: Image Communication*, 30:37–56, 2015. 3

[23] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proc. of Computer Vision and Pattern Recognition*, pages 1651–1660, 2020. 3

[24] Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Transactions on Image Processing*, 26(5):2519–2532, 2017. 3

[25] Stephen Mangiat and Jerry Gibson. High dynamic range video with ghost removal. In *Proc. of Applications of Digital Image Processing*, pages 307–314. SPIE, 2010. 3

[26] Stephen Mangiat and Jerry Gibson. Spatially adaptive filtering for registration artifact removal in HDR video. In *Proc. of International Conference on Image Processing*, pages 1317–1320, 2011. 3

[27] Rafal K Mantiuk, Dounia Hammou, and Param Hanji. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625*, 2023. 8

[28] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Proc. of Pacific Conference on Computer Graphics and Applications*, pages 382–390, 2007. 3

[29] Srinivasa G Narasimhan and Shree K Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):518–530, 2005. 1, 3

[30] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet. HDR-VQM: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46–60, 2015. 8

[31] Nayar and Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1168–1175. IEEE, 2003. 2, 3

[32] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. of Computer Vision and Pattern Recognition*, pages 472–479, 2000. 1, 3

[33] Yutaro Okamoto, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi. Deep snapshot HDR imaging using multi-exposure color filter array. *The Visual Computer*, 40 (5):3285–3301, 2024. 1, 3

[34] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Proc. of Advances in Neural Information Processing Systems*. 2019. 1

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6, 1

[36] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):203–1, 2012. 3

[37] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proc. of Computer Vision and Pattern Recognition*, 2017. 5

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia

Polosukhin. Attention is all you need. *Proc. of Advances in Neural Information Processing Systems*, 30, 2017. 6, 1

[39] Hongcheng Wang, Ramesh Raskar, and Narendra Ahuja. High dynamic range video using split aperture camera. In *IEEE 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Washington, DC, USA*. Citeseer, 2005. 2, 3

[40] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 8

[41] Qingsen Yan, Lei Zhang, Yu Liu, Yu Zhu, Jinqiu Sun, Qinfeng Shi, and Yanning Zhang. Deep HDR imaging via a non-local network. *IEEE Transactions on Image Processing*, 29: 4308–4322, 2020. 3

[42] Siqi Yang, Zhaojun Huang, Yakun Chang, Bin Fan, Zhaofei Yu, and Boxin Shi. Real-data-driven 2000 FPS color video from mosaicked chromatic spikes. In *Proc. of European Conference on Computer Vision*, pages 1111–2222, 2024. 3

[43] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proc. of Computer Vision and Pattern Recognition*, pages 11996–12005, 2021. 3

[44] Jing Zhao, Ruiqin Xiong, Jiyu Xie, Boxin Shi, Zhaofei Yu, Wen Gao, and Tiejun Huang. Reconstructing clear image for high-speed motion scene with a retina-inspired spike camera. *IEEE Transactions on Computational Imaging*, 8:12–27, 2021. 1

[45] Jing Zhao, Ruiqin Xiong, Jian Zhang, Rui Zhao, Hangfan Liu, and Tiejun Huang. Learning to super-resolve dynamic scenes for neuromorphic spike camera. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 3579–3587, 2023. 3

[46] Junwei Zhao, Jianming Ye, Shiliang Shiliang, Zhaofei Yu, and Tiejun Huang. Recognizing high-speed moving objects with spike camera. In *Proc. of ACM MM*, pages 7657–7665, 2023. 3

[47] Rui Zhao, Ruiqin Xiong, Jian Zhang, Zhaofei Yu, Shuyuan Zhu, Lei Ma, and Tiejun Huang. Spike camera image reconstruction using deep spiking neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(6): 5207–5212, 2023. 1

[48] Yajing Zheng, Lingxiao Zheng, Zhaofei Yu, Boxin Shi, Yonghong Tian, and Tiejun Huang. High-speed image reconstruction through short-term plasticity for spiking cameras. In *Proc. of Computer Vision and Pattern Recognition*, pages 6358–6367, 2021. 1

[49] Lin Zhu, Siwei Dong, Tiejun Huang, and Yonghong Tian. A retina-inspired sampling method for visual texture reconstruction. In *Proc. of International Conference on Multimedia and Expo*, pages 1432–1437, 2019. 1, 3

[50] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via spiking neural model. In *Proc. of Computer Vision and Pattern Recognition*, pages 1438–1446, 2020. 1, 3

[51] Zhenkun Zhu, Ruiqin Xiong, Jing Zhao, Rui Zhao, Xiaopeng Fan, Shuyuan Zhu, and Tiejun Huang. High dynamic range imaging for dynamic scenes based on multi-level spike camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025. 1