The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Robust Noise Modeling for Spike Camera via Time-Interval Quantification and Spike-DSLR Multimodal Dataset in Low-Light Imaging

Yue Cao[1], **Sizhao Li**[1], **Liguo Zhang**[1]*
[1]College of Computer Science and Technology, Harbin Engineering University, Heilongjiang, China
{cscaoyue, sizhao.li, zhangliguo}@hrbeu.edu.cn

Abstract

The inherent differences between spike cameras and traditional frame-based cameras lead to more complex and diverse noise characteristics, particularly under extremely low-light conditions. Existing noise modeling approaches for spike camera predominantly rely on inter-spike intervals (ISI) for noise quantification, which often results in inaccurate noise characterization. Moreover, current datasets for spike camera image reconstruction tasks are either synthetic or lack corresponding high-quality reference images, severely limiting rigorous evaluation of noise modeling methods. To address this limitation, we propose a multimodal noise modeling framework for spike camera that integrates insights from traditional frame-based imaging into spike imaging. Specifically, we introduce a time-interval-based quantification method inspired by the exposure-time concept used in traditional frame-based cameras, enabling accurate noise characterization for spike camera. Furthermore, we present the Spike-DSLR Multimodal Dataset (SDMD), the first real-world dataset capturing aligned multimodal data pairs from spike cameras and Digital Single-Lens Reflex (DSLR) cameras, explicitly designed for evaluating spike camera noise models. Experimental results on SDMD demonstrate that our noise modeling approach significantly enhances spike camera image reconstruction quality under low-light conditions, achieving more than 1.6 dB improvement in PSNR compared to existing state-of-the-art methods. This validates both the necessity and effectiveness of adopting a multimodal perspective in spike camera noise modeling.

Introduction

With the continuous advancement of imaging technology, research on low-light noise modeling and denoising methods has emerged as a prominent topic in computer vision. Traditional frame-based cameras exhibit inherent limitations in capturing dynamic light variations during exposure periods: they cannot record detailed temporal changes within the exposure interval. Even video recording remains constrained by finite frame rates, resulting in information loss between consecutive exposures.

To address these limitations and further advance imaging technology, Peking University introduced the Spike Vision Model in 2015, replacing the conventional concept of "Video" with "Vidar" (visual radar). This model effectively compensates for the limitations of traditional frame-based cameras in high-speed scene recording by continuously capturing dynamic light variations, thereby offering a novel approach to light information acquisition. Spike cameras are inspired by biological retinas, applying event-driven encoding principles where visual information is represented through discrete spikes rather than continuous signals. This innovation demonstrates significant potential not only in visual information representation but also exerts profound influence on the study of biological mechanisms. With their high temporal resolution and sensitivity to light intensity variations, spike cameras have made significant advances in reconstruction applications, particularly in high dynamic range imaging (Chang et al. 2023, 2024; Han et al. 2020) and image deblurring (Chen et al. 2024).

Spike cameras represent a revolutionary sensing technology that captures visual information through asynchronous binary data rather than traditional frame-based exposure. While promising significant advantages in temporal resolution and dynamic range, these cameras face substantial challenges in extremely low-light environments where the inherent sparsity of spike data and complex noise characteristics severely impact image reconstruction quality. Addressing these challenges requires a multimodal perspective that bridges the conceptual frameworks of traditional imaging and neuromorphic sensing.

Current approaches to spike camera image reconstruction predominantly rely on simulated data, as capturing paired ground truth references for real spike data remains technically challenging. However, existing simulation methods are primarily designed for normal or moderately low-light scenarios and fail to accurately model the noise characteristic under extremely low-light conditions. This disconnect between simulation and real-world performance highlights the need for both improved noise modeling techniques and high-quality multimodal datasets that enable reliable evaluation.

To address these limitations, we propose a time-interval-based noise modeling approach that fundamentally rethinks how spike camera noise should be quantified. Existing methods rely on Inter-Spike Interval (ISI) modeling, which inherently couples noise with light intensity and overlooks critical temporal aspects of spike data simulation. Our ap-

---

*Corresponding author.

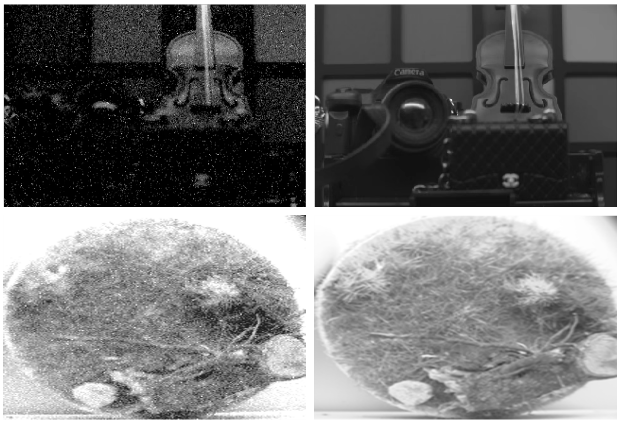
TFP result from spike camera Reference by DSLR camera

Figure 1: Visualization of paired data from the SDMD dataset under extreme low-light conditions (illuminance < 5 lux).

proach draws inspiration from frame-based camera noise characterization, treating sampling time as a sequence of discrete intervals within which various noise types are meticulously modeled. This cross-modal knowledge transfer enables more accurate noise parameter estimation by ensuring that dark current noise remains independent of light intensity and properly accounts for charging time effects considerations that traditional ISI-based noise modeling approaches cannot address.

Moreover, we introduce the Spike-DSLR Multimodal Dataset (SDMD), a novel dataset that pairs real spike camera data with corresponding high-quality reference images captured by Digital Single-Lens Reflex (DSLR) cameras (as shown in Fig. 1). This multimodal approach provides, for the first time, a reliable benchmark for evaluating different spike simulation methods against ground truth data, enabling objective assessment of noise modeling techniques. Without this cross-modal reference framework, accurately evaluating spike camera noise models would remain virtually impossible.

To sum up, the main contributions of this work include:

- We conduct an in-depth investigation of the spike camera imaging process and propose a spike camera noise model based on time interval modeling. This model comprehensively considers the characteristics of spike cameras, typical noise types, and spike camera image reconstruction algorithms, thereby more completely capturing and describing complex noise characteristics under extremely low-light conditions.
- We establish a Spike-DSLR Multimodal Dataset (SDMD) specifically designed for extremely low-light scenarios, containing paired noisy spike data and high-quality reference image irradiance, providing a crucial data foundation for future research in noise modeling and spike camera image reconstruction.
- Experimental results demonstrate that our proposed method can generate spike data distributions that accurately reflect real-world extremely low-light scenarios, significantly enhancing spike camera image reconstruction performance under these challenging conditions.

Ralated Work

Image Reconstruction for Spike Cameras

The core task of spike cameras is image reconstruction, which involves mapping from spike camera data to image irradiance. The development of this task can be categorized into two primary phases: traditional reconstruction methods and deep learning-based approaches. Among traditional methods, Texture from Playback (TFP) and Texture from Inter-Spike Interval (TFI) (Zhu et al. 2019) reconstruct image irradiance by analyzing spike counts and intervals. TFP accumulates spike data through temporal windows, adapting to various dynamic ranges and contrasts, while TFI specializes in capturing contours in high-speed motion scenarios.

With the rapid development of deep learning technologies, spike camera image reconstruction methods have achieved significant breakthroughs, further enhancing reconstruction quality and accuracy. Spk2ImgNet (Zhao et al. 2021) leverages deep learning by integrating local and long-term aligned temporal information, significantly improving reconstruction results. SSML (Chen et al. 2022) employs a self-supervised learning framework, combining blind spot networks and pseudo-labels to further enhance spike camera image reconstruction precision. Concurrently, BSF (Zhao et al. 2024b) improves reconstruction robustness and accuracy through higher-order spike emission time differential analysis, quantization fluctuation modeling, and multi-scale feature alignment techniques.

Moreover, researchers have developed specialized algorithms for specific spike camera application scenarios. NeuroZoom (Duan et al. 2023) addresses denoising and super-resolution challenges for neuromorphic cameras, substantially improving image quality for dynamic vision sensors and spike cameras. SSIR (Zhao et al. 2024a) achieves reconstruction performance comparable to state-of-the-art methods while reducing computational requirements through energy-efficient spiking neural networks. The advancement of these methodologies has not only driven progress in spike camera technology but also provided robust support for future applications.

Sensor Noise Modeling

Research in sensor noise modeling for traditional frame-based camera has evolved from physics-based statistical models to data-driven approaches. Early sensor noise modeling methods characterized complex noise components through parameterized approaches such as Poisson-Gaussian mixture distributions (Foi et al. 2008; Foi 2009). Related studies (Wei et al. 2021; Zhang et al. 2021; Feng et al. 2024; Wang et al. 2020) inferred camera gain by capturing data under uniform illumination and applying photon transfer techniques (Janesick, Klaasen, and Elliott 1985) to model shot noise. For sensor-generated noise, ELD (Wei et al. 2021) modeled the heavy-tailed characteristics of dark current noise using Tukey Lambda distributions (Joiner and

Rosenblatt 1971), ELLE (Wang et al. 2021) employed uniform distributions to model black level noise, while PMN (Feng et al. 2024) adopted a decomposition strategy for dark current fixed pattern noise.

In recent years, data-driven models have become a focal point in noise modeling research. NoiseFlow (Abdelhamed, Brubaker, and Brown 2019) modeled noise distributions through normalizing flows, while CA-GAN (Chang et al. 2020) introduced Generative Adversarial Networks (GANs) to enhance noise realism. However, these methods primarily target normal or low-light conditions (illuminance greater than 5 lux), limiting their applicability in extreme low-light scenarios. To overcome this limitation, LRD (Zhang et al. 2023) proposed a generalized low-light rawRGB noise synthesis and modeling method that effectively enhances noise modeling capabilities in complex scenes by simulating noise characteristics in low-light environments. This reveals the current limitations of data-driven models in extremely dark scenes, which, due to neglecting the physical noise generation process, underperform compared to traditional physics-based statistical models (Zhang et al. 2021; Wei et al. 2021). Furthermore, LLD (Cao et al. 2023) significantly improved noise modeling precision by designing a learnable noise modeling network structure and fully considering the mapping relationship between ISO and noise levels.

Due to differences in sampling mechanisms between spike cameras and traditional frame-based cameras, noise modeling for this emerging imaging technology has attracted extensive attention. Existing noise modeling methods for spike camera (Zhu et al. 2023, 2021) divide sampling time into multiple discharge periods, each corresponding to an inter-spike interval, first generating inter-spike intervals and then converting them to spike data. Specifically, NeuSpike-Net (Zhu et al. 2021) extended the capability to handle dark current fixed pattern noise in spike cameras but primarily focused on normal lighting conditions. RSIR (Zhu et al. 2023), through physics-based statistical noise modeling, made the first attempt to reconstruct high-quality image irradiance under low-light conditions. However, in extreme low-light scenarios, these methods still demonstrate limitations.

High-Quality Real Noise Datasets Construction

Imaging technology has evolved toward multi-device, multi-scene directions, driving data collection paradigms from traditional frame-based cameras to emerging camera domains, and from normal light to extremely low-light scenes.

Data Collection for Traditional Frame Cameras: Early research primarily constructed paired data by controlling ISO and exposure time. Nam et al.(Nam et al. 2016) built cross-channel noise models using multi-ISO static scene captures, but failed to consider the effects of dynamic lighting changes; while the RENOIR dataset(Anaya and Barbu 2018) expanded exposure parameter combinations, spatial misalignment errors resulted from neglecting dark current noise calibration. These datasets reveal the domain gap problem between controlled laboratory conditions and complex real-world scenarios.

Data Collection for Low-Light Scenes: SIDD (Abdelhamed, Lin, and Brown 2018) improved data fidelity by averaging multiple frames to generate reference images, but faced alignment sensitivity and fixed pattern noise residual issues in mobile device imaging. For extremely dark scenes, Chen et al. (Chen et al. 2018) pioneered a method pairing long and short exposures based on extremely low-light rawRGB data. Although DND (Plotz and Roth 2017) and ELD (Wei et al. 2021) constructed data through high-low ISO combinations, they were limited to use as test sets due to the limited number of paired data.

Data Collection for New Imaging Devices: LED (Duan 2024) captured background active noise events using a dual-camera system, but the threshold-driven mechanism of event cameras fundamentally differs from the photon statistical models of traditional cameras; the RSIR spike dataset only provides low-light noise data, lacking noise correlation analysis with traditional frame-based devices. The differences in data collection paradigms between emerging and traditional cameras highlight the theoretical challenges of cross-domain modeling.

Methodology

Working Principles and Imaging Mechanisms of Spike Cameras

The defining feature of spike cameras is their asynchronously operating pixel structure, enabling each pixel to independently sense light intensity changes with fine-grained temporal resolution. The imaging process of a spike camera primarily comprises three states: integration, reset, and readout.

During integration, the photodiode accumulates photons, generating a photocurrent I_{ph} that decreases voltage across capacitor C_{pd}. When this voltage reaches the reference voltage V_{ref}, the comparator triggers. In the reset state, the system clears accumulated charges, resets to supply voltage V_{dd}, and generates a 1-bit signal. Each pixel independently produces a spike stream encoded as a binary sequence, where '1' represents a spike trigger and '0' indicates ongoing accumulation. The Inter-Spike Interval (ISI), the duration between spikes, inversely correlates with light intensity. The readout phase then transmits this signal to the data bus.

The threshold voltage θ represents the difference between V_{dd} and V_{ref}, while V_{ph} denotes the photovoltage. Spike cameras output spatio-temporal binary streams with dimensions $H \times W \times M$, where spatial resolution is $H \times W$ and M represents the number of time intervals T_r within total sampling time T. The TFP algorithm reconstructs image irradiance by calculating $\mathrm{TFP}(\hat{S}) = \frac{\sum_{m=1}^{M} \hat{S}(m)}{M}$, where pixel positions are indexed by (i, j) and time intervals by m.

Problem Formulation

To establish the more comprehensive spike camera noise model by integrating spike camera imaging characteristics and addressing spike camera image reconstruction requirements, two key issues must be addressed: **algorithm input** and **noise quantification method**.

Algorithm 1: Spike simulation algorithm under ideal conditions without shot and dark current noise.

```
**Input**: Image irradiance $I(t)$ varying with time $t$, total sampling time $T = M \cdot T_r$
**Output**: Synthesized spike stream $\hat{S} = \{\hat{S}(i,j,m)\}_{m=1,\dots,M}$
**Initialization**: $\hat{S}(i,j,m) = 0$
**for** $m$ in range $[1, M]$ **do**
    Update voltage: $V_{ph}(i,j) \leftarrow V_{ph}(i,j) + \frac{1}{C_{pd}} \cdot I(i,j,m \cdot T_r) \cdot T_r$
    **if** $V_{ph}(i,j) \geq \theta$ **then**
        Generate spike signal: $\hat{S}(i,j,m) \leftarrow 1$
        Reset voltage: $V_{ph}(i,j,m) \leftarrow 0$
    **end if**
**end for**
```

Algorithm Input: For spike data simulation algorithm, input data can be either clean spike streams or continuous clean image sequences. One approach to obtain clean spike streams is to assume the absence of shot noise and dark current noise. Under these idealized conditions, ideal spike streams could be generated through the spike camera imaging process. Drawing on the SpikingSIM (Zhao et al. 2022) method, Algorithm 1 summarizes the spike camera simulation process under ideal conditions without shot noise and dark current noise. However, even under these assumptions, clean spike streams remain practically unobtainable, primarily due to unavoidable quantization noise. Since the photovoltage V_{ph} cannot precisely equal the threshold voltage θ at a specific time interval point, V_{ph} inevitably incurs quantization errors. Furthermore, these ideal conditions without shot noise and dark current noise are unachievable in reality. Consequently, clean spike streams do not exist and cannot serve as algorithm input. Based on this analysis, we use continuous clean image sequences as input for spike data simulation. The generated spike data is then used as input for reconstruction algorithms, which output continuous image sequences.

Noise Quantification Method: Having considered appropriate algorithm input, we now examine noise quantification methods in the spike camera imaging process. Existing spike data simulation methods for low-light scenes divide sampling time into multiple discharge periods, each corresponding to an inter-spike interval, first generating inter-spike intervals and then converting them to spike data. While NeuSpike-Net extended capability to handle dark current fixed pattern noise and RSIR made the first attempt to reconstruct high-quality image irradiance under low-light conditions through physics-based statistical noise modeling, these approaches still demonstrate significant limitations in extreme low-light scenarios.

The fundamental problem with inter-spike interval models lies in their inherent coupling of dark current noise with light intensity, making it impossible to isolate and analyze dark current noise characteristics independently. This forces researchers to use uniform illumination data to infer dark current noise parameters, resulting in imprecise modeling. Additionally, these models ignore charging time periods when the sensor doesn't receive light signals. The integer form of inter-spike intervals further lacks the precision needed for the refined noise characterization required in extreme low-light conditions.

To overcome these challenges, we propose a time-interval-based noise quantification method that treats sampling time as a sequence of intervals, modeling various noise types within each interval. This approach draws inspiration from traditional frame-based camera noise parameter estimation techniques, ensuring that dark current noise in individual time intervals remains unaffected by light intensity, thereby enhancing the accuracy of noise parameter estimation while effectively accounting for charging time effects. Our method provides a more comprehensive framework for characterizing spike camera noise in extreme low-light conditions.

Noise Analysis and Modeling Methods for Spike Cameras

As shown in Fig. 2, we propose a comprehensive noise modeling method specifically designed for spike cameras in extremely low-light conditions. This approach systematically analyzes each noise component and develops a Time-Interval-based Spike Simulation method for Extreme Low-Light Conditions (ELSSim). The method comprehensively considers multiple noise sources, including shot noise N_{shot}^{spike}, dark current noise N_{DC}^{spike}, dark current hot-pixel noise N_{HP}^{spike}, photo-response non-uniformity noise N_{PRNU}^{spike}, and refractory period noise N_{RP}^{spike}, aiming to precisely characterize the noise behavior of spike cameras in extremely low-light scenes.

Inspired by physical statistical noise models, our method leverages the analysis of Flat-field Frames (FF) and Dark Frames (DF) to support noise modeling. Specifically, we collected 20 sets of spike data under completely dark conditions (denoted as S_{DF}^{k1}, where $k1 = 1, 2, \dots, 20$) for statistical analysis of dark current noise characteristics. To simulate noise under different lighting conditions, we further collected flat-field spike data under various brightness conditions without a lens (denoted as S_{FF}^{k2}, where $k2 = 1, 2, \dots, 20$).

- **Shot Noise** Shot noise N_{shot}^{spike} originates from the quantum nature of photons and is an unavoidable phenomenon at the physical level (Konnik and Welsh 2014; Gow et al. 2007), with significant impact on imaging quality. The shot noise can be directly modeled using the Poisson distribution $\mathcal{P}$, effectively avoiding the complex gain correction processes common in traditional frame-based camera noise modeling.

- **Dark Current Noise** Referencing the research in LLD (Cao et al. 2023), dark current noise can be divided into two main components: fixed pattern noise and shot noise. To adapt to the time-interval-based modeling strategy, our method defines dark current noise within a single time interval. Specifically, dark current noise N_{DC}^{spike} can be modeled using the Poisson distribution $\mathcal{P}$ as follows:

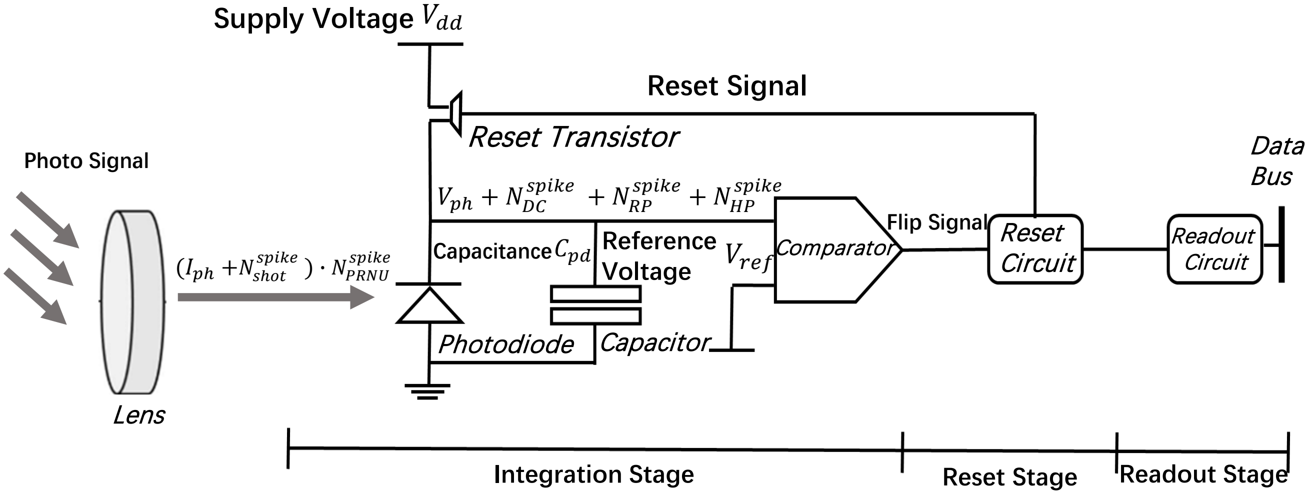

Figure 2: Pipeline of the spike camera in real-world scenarios.

$$N_{DC}^{spike} \sim \mathcal{P}\left(\mu_{DC}\right), \tag{1}$$

where the dark current fixed pattern noise μ_{DC} for a single time interval is based on the statistical results of 20 sets of dark-field spike data, calculated as:

$$\mu_{DC}(i,j) = \frac{\sum_{k1=1}^{20}\sum_{m=1}^{M} S_{DF}^{k1}(i,j)}{20 \cdot M}, \tag{2}$$

This method effectively reduces noise fluctuations in measurements by averaging multiple sets of dark-field spike data, more accurately reflecting the fixed pattern noise differences between pixels.

- **Dark Current Hot-Pixel Noise** The model for dark current hot-pixel noise can be constructed by setting a hot-pixel threshold θ_{HT} to determine the hot-pixel location matrix $Mask(i,j)$:

$$Mask(i,j) = \begin{cases} 1, & (\mathrm{TFP}(S_{DF}^{k1=1}(i,j)) - \mu_{DC}(i,j)) \geq \theta_{HT} \\ 0, & (\mathrm{TFP}(S_{DF}^{k1=1}(i,j)) - \mu_{DC}(i,j)) < \theta_{HT} \end{cases}. \tag{3}$$

Once the hot-pixel locations are determined, dark current hot-pixel noise N_{HP}^{spike} can be modeled for these specific positions to more accurately reflect the noise characteristics:

$$N_{HP}^{spike} \sim \mathcal{N}\left(\mu_{HP}, \sigma_{HP}^{2}\right), \tag{4}$$

In this model, the mean μ_{HP} and noise intensity σ_{HP} can be derived through the operation $Mask \times (\mathrm{TFP}(S_{DF}^{k1}) - \mu_{DC})$, further quantifying the characteristics of hot-pixel noise.

- **Photo-Response Non-Uniformity Noise** Photo-Response Non-Uniformity (PRNU) noise is a fixed pattern noise caused by non-uniformities in the sensor manufacturing process, which significantly impacts data quality during imaging. Specifically, even under the same lighting conditions, due to manufacturing non-uniformities, each pixel responds slightly differently to light, resulting in PRNU noise manifesting as non-uniform light response distributions during imaging. PRNU noise parameters can be calibrated and modeled by obtaining flat-field spike data through lens-less shooting, thus mitigating its impact on imaging quality.

To model 20 sets of flat-field spike data S_{FF}^{k2} under different brightness conditions, we first calculate the average response value $\overline{S_{FF}^{k2}}$ for each set to provide a baseline for subsequent PRNU noise analysis. The average response value is calculated as:

$$\overline{S_{FF}^{k2}} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\mathrm{TFP}(S_{FF}^{k2}) - \mu_{DC}\right)(i,j)}{H \cdot W}, \tag{5}$$

Next, PRNU noise N_{PRNU}^{spike} is defined as a tensor of size $H \times W$, which helps systematically represent and analyze the photo-response non-uniformity of each pixel. $N_{PRNU}^{spike}(i,j)$ represents the slope parameter obtained through linear regression between the dark-current-corrected TFP values $(\mathrm{TFP}(S_{FF}^{k2}) - \mu_{DC})(i,j)$. $N_{PRNU}^{spike}(i,j)$ and the average response value $\overline{S_{FF}^{k2}}$ at pixel position (i,j).

- **Refractory Period Noise** In neuroscience, the refractory period (RP) refers to the time interval during which a neuron, after experiencing an action potential, temporarily cannot generate subsequent action potentials, a phenomenon of significant importance in neural signal transmission. In bio-inspired imaging systems such as spike cameras and event cameras, when a spike or event is generated, the relevant pixel temporarily loses its response capability for a certain period, thus defining this time interval as the refractory period (Low and Lee 2023), a definition inspired by concepts from neuroscience and physiology.In traditional frame-based cameras, the discharge time of the photodiode

Algorithm 2: Spike stream simulation algorithm in extreme low-light scenarios.

```
Input: Image irradiance $I(t)$ varying with time $t$, total sampling time $T = M \cdot T_r$
Output: Synthesized spike stream $\hat{S} = \hat{S}(i, j, m)_{m=1,\ldots,M}$
Initialization: $\hat{S}(i, j, m) = 0$
for $m$ in range $[1, M]$ do
    Add shot noise and photo-response non-uniformity noise
    $I(i, j, m \cdot T_r) \leftarrow \mathcal{P}(I(i, j, m \cdot T_r)) \cdot N_{PRNU}^{spike}$
    Add dark current noise and refractory period noise
    $V_{ph}(i, j) \leftarrow V_{ph}(i, j) + \frac{1}{C_{pd}} \cdot I(i, j, m \cdot T_r) \cdot \delta t \cdot \left(\frac{T_r}{\delta t} + N_{RP}^{spike}\right) + N_{DC}^{spike} \cdot \theta$
    Add dark current hot-pixel noise
    $\text{Mask}(i, j) \cdot V_{ph}(i, j) \leftarrow \text{Mask}(i, j) \cdot V_{ph}(i, j) + N_{HP}^{spike} \cdot \theta$
    if $V_{ph}(i, j) \geq \theta$ then
        Generate spike signal: $\hat{S}(i, j, m) \leftarrow 1$
        Reset voltage: $V_{ph}(i, j, m) \leftarrow 0$
    end if
end for
```

corresponds to the exposure time, while the charging process is completed during the interval between two shutters. In spike cameras, the time interval T_r is subdivided into several time units δt, and the charging operation is completed within one time unit after spike triggering, which is considered the refractory period, during which the photodiode completes charging. The operating clock frequency is $CLK = 10$ MHz, so the time unit δt can be calculated as $\delta t = \frac{2}{CLK}$ seconds, resulting in $\frac{T_r}{\delta t}$ time units within one time interval T_r. In spike cameras, the photodiode discharge time covers the entire integration cycle, while the charging process is completed within one time unit δt after spike generation, a relationship that significantly impacts the imaging process. During the refractory period, light signals cannot be effectively collected, thus this loss of light signal is defined as refractory period noise N_{RP}^{spike}, a typical device-related noise whose presence may significantly affect the overall performance of the imaging system. Based on the above analysis, considering the impact of refractory periods on light signal collection, refractory period noise N_{RP}^{spike} can be defined as:

$$N_{RP}^{spike}(i, j) = \begin{cases} -1, & \hat{S}(i, j, m-1) = 1 \\ 0, & \hat{S}(i, j, m-1) = 0 \end{cases} \quad \text{when } m > 1 \ . \tag{6}$$

•Spike Data Synthesis Algorithm for Extremely Low-Light Scenes Based on the time-interval model description and the definition and modeling of various noise types discussed above, we can derive a spike data simulation algorithm suitable for extremely low-light scenes. As shown in Algorithm 2, the entire synthesis process strictly follows the actual imaging process of spike cameras. Specifically, dark current hot-pixel noise reflects the inherent noise characteristics of the sensor under no illumination conditions, while refractory period noise simulates the brief non-responsive period of spike camera pixels after triggering a spike. The inclusion of these two noise types makes our model more closely aligned with the physical characteristics of actual spike cameras.

Spike-DSLR Multimodal Dataset in Extreme Low-Light Conditions

This section introduces a novel Spike-DSLR Multimodal Dataset (SDMD). The construction methodology aligns with the spike camera image reconstruction task, which aims to reconstruct image irradiance from spike data. We constructed multimodal system of spike camera and DSLR on the optical platform. We carefully controlled the light box illuminance to be **below 5 lux** to capture images under extreme low-light environments, consistent with the standard definition established in previous literature (Chen et al. 2018; Wei et al. 2021).

The entire construction process of the Spike-DSLR Multimodal Dataset consists of the following seven steps: (1) we captured images with an exposure time of $1/13$ second and ISO setting of 100, obtaining images with relatively high signal-to-noise ratios. (2) To further reduce shot noise, dark current shot noise, and row noise in the images, we adopted a 20-frames averaging strategy, referencing the SIDD dataset (Abdelhamed, Lin, and Brown 2018). (3) To address fixed pattern noise, including dark current fixed pattern noise and PRNU noise, we followed the PMN approach(Feng et al. 2024) and captured 400 dark-field images under completely black conditions using identical exposure parameters. Through multi-frame averaging, we extracted the combination of dark current fixed pattern noise and black level error noise. (4) We calibrated PRNU noise using flat-field images captured by the DSLR camera. By removing these two types of fixed pattern noise, we significantly eliminated noise in the reference images, ensuring their clarity and reliability. (5) We adopted the existing pixel-shift shooting technology (Zhang, Fu, and Li 2022; Qian et al. 2022) to avoid demosaicing operations, thereby obtaining single-channel clean images and further improving dataset quality. (6) We determined the lens shading correction matrix and camera response function for the DSLR camera, applying further corrections to the clean images. (7) Finally, to ensure precise alignment between DSLR and spike camera modalities, we adopt the same method as NeuSpike-Net to use homography matrices between single-channel clean images and TFP-reconstructed images for registration, generating high-quality reference images. Through this carefully crafted series of steps, we successfully constructed a multimodal dataset that combines authenticity with high quality, providing a reliable foundation for subsequent spike camera noise research. Ultimately, we captured paired multimodal data for 100 different scenes.

Experimental Results and Analysis

Experimental Settings and Implementation Details

This section consists of two parts: the experimental setup and the dataset setup.

•Experimental Setup

This section compares four classical spike camera image reconstruction methods: TFP, TFI, Spk2ImgNet, SSML and BSF. These methods were all simulated under normal lighting conditions and their spike algorithms were constructed accordingly. However, the RSIR experiments demonstrate that these methods perform poorly in low-light scenarios. To address this issue, RSIR proposed a specialized noise model and corresponding spike camera image reconstruction network for low-light scenes. As environmental conditions deteriorate further to extreme low-light scenarios, we introduces a novel spike camera noise modeling and constructs the SDMD dataset comprising spike camera and DSLR camera captures. To ensure experimental fairness, we employ the same spike camera image reconstruction network architecture and training settings as the RSIR method, differing only in the simulation method, which utilizes our proposed ELSSim algorithm. The RSIR method preprocesses raw spike data using a sampling quantity of $M = 320$ and TFP processing with a window size of 32, ultimately generating $H \times W \times 10$ preprocessed data as input to the reconstruction network. Our method adopts the same preprocessing settings as RSIR. Additionally, ”Paired Data” refers to the RSIR spike camera image reconstruction network trained using the SDMD training set.

• Dataset Setup

Simulated Spike Dataset: To maintain consistency with the RSIR method’s training setup, the simulated spike dataset utilizes the DAVIS(Perazzi et al. 2016) dataset, encompassing 90 distinct scenes. Among these, 85 scenes were used for synthesizing training data, while the remaining 5 sets of scenes were used for synthesizing test data. The test spike data was derived from these 5 sets of clear images synthesized with Algorithm 2.

SDMD Real Spike Dataset: The SDMD dataset comprises 100 different scenes, with 20 sets designated as the test set for evaluating the performance of various spike camera image reconstruction methods, and the remaining 80 sets serving as SDMD training data, primarily for training the ”Paired Data” experiment. The primary purpose of the SDMD dataset is to verify the image reconstruction capabilities of reconstruction models trained based on different spike simulation algorithms, rather than providing a training scheme based on real paired datasets.

Extreme Low-light Dynamic Spike Dataset: We present ELDSD (Extreme Low-light Dynamic Spike Dataset), a specialized collection of spike camera recordings captured under challenging illumination conditions. This dataset comprises 15 diverse dynamic scenes recorded in both indoor and outdoor environments, with illuminance levels meticulously controlled below 5 lux.

RSIR Real Spike Dataset: The RSIR dataset contains 10 different scenes, all used to evaluate the performance of different spike camera image reconstruction methods. Unlike the SDMD dataset, the RSIR dataset does not provide reference images, thus quality comparisons of reconstructed images can only be made through subjective human visual assessment.

Methods	PSNR	SSIM
TFP($M = 320$)	13.02	0.146
TFP($M = 4000$)	13.45	0.266
TFI	13.62	0.293
Spk2ImgNet	13.73	0.314
SSML	15.70	0.397
BSF	15.72	0.401
RSIR	16.20	0.438
Paired Data	18.03	0.662
ELSSim	17.81	0.681

Table 1: The PSNR(dB) and SSIM results of different methods on SDMD dataset.

Comparison of Experimental Results

As shown in Table. 1, the classical TFI method relies on the fundamental correspondence between spike signals and image irradiance to reconstruct images. In contrast, the Spk2ImgNet and SSML methods are optimized solely for high-speed motion scenes under normal illumination, resulting in limited applicability under more complex lighting conditions. When confronted with complex scene noise, these methods struggle to meet the requirements of a wide range of applications. The RSIR method achieves certain performance improvements by considering factors such as fixed pattern noise. However, RSIR employs a ISI-based modeling approach that estimates noise intensity within a single spike interval, resulting in a correlation between light intensity and dark current noise intensity, thereby limiting the extent of its performance improvement. In comparison, our proposed method adopts a time-interval-based noise modeling approach and demonstrates excellent performance in extreme low-light scenarios.

Conclusion

In this paper, we advance neuromorphic visual sensing through our multimodal approach to spike camera noise modeling in extreme low-light conditions. Our time-interval-based noise quantification method overcomes limitations of traditional techniques by drawing inspiration from frame-based camera characterization. Complementing this, our novel Spike-DSLR Multimodal Dataset (SDMD) pairs real spike data with DSLR reference images, establishing a benchmark for validating noise models against ground truth.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 62502112, National Key Research and Development Program of Heilongjiang Province under Grant No. 2024ZXDXA12, and the Fundamental Research funds for the Central Universities under Grant No. 3072025YC0602.

References

Abdelhamed, A.; Brubaker, M. A.; and Brown, M. S. 2019. Noise Flow: Noise modeling with conditional normalizing flows. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 3165–3173.

Abdelhamed, A.; Lin, S.; and Brown, M. S. 2018. A high-quality denoising dataset for smartphone cameras. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1692–1700.

Anaya, J.; and Barbu, A. 2018. RENOIR-a benchmark dataset for real noise reduction evaluation. *Journal of Visual Communication and Image Representation (JVCI)*, 51: 144–154.

Cao, Y.; Liu, M.; Liu, S.; Wang, X.; Lei, L.; and Zuo, W. 2023. Physics-Guided ISO-Dependent Sensor Noise Modeling for Extreme Low-Light Photography. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5744–5753.

Chang, K.-C.; Wang, R.; Lin, H.-J.; Liu, Y.-L.; Chen, C.-P.; Chang, Y.-L.; and Chen, H.-T. 2020. Learning camera-aware noise models. In *European Conference on Computer Vision (ECCV)*, 343–358.

Chang, Y.; Xiaokaiti, Y.; Liu, Y.; Fan, B.; Huang, Z.; Huang, T.; and Shi, B. 2024. Towards HDR and HFR Video from Rolling-Mixed-Bit Spikings. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25117–25127.

Chang, Y.; Zhou, C.; Hong, Y.; Hu, L.; Xu, C.; Huang, T.; and Shi, B. 2023. 1000 FPS HDR video with a spike-RGB hybrid camera. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22180–22190.

Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to see in the dark. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3291–3300.

Chen, S.; Duan, C.; Yu, Z.; Xiong, R.; and Huang, T. 2022. Self-Supervised Mutual Learning for Dynamic Scene Reconstruction of Spiking Camera. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2859–2866.

Chen, S.; Zhang, J.; Zheng, Y.; Huang, T.; and Yu, Z. 2024. Enhancing motion deblurring in high-speed scenes with spike streams. *Advances in Neural Information Processing Systems (NeurIPS)*, 70279–70292.

Duan, P.; Ma, Y.; Zhou, X.; Shi, X.; Wang, Z. W.; Huang, T.; and Shi, B. 2023. NeuroZoom: Denoising and Super Resolving Neuromorphic Events and Spikes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 45(12): 15219–15232.

Duan, Y. 2024. LED: A Large-scale Real-world Paired Dataset for Event Camera Denoising. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25637–25647.

Feng, H.; Wang, L.; Wang, Y.; Fan, H.; and Huang, H. 2024. Learnability Enhancement for Low-Light Raw Image Denoising: A Data Perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 46(1): 370–387.

Foi, A. 2009. Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 89(12): 2609–2629.

Foi, A.; Trimeche, M.; Katkovnik, V.; and Egiazarian, K. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing (T-IP)*, 17(10): 1737–1754.

Gow, R. D.; Renshaw, D.; Findlater, K.; Grant, L.; McLeod, S. J.; Hart, J.; and Nicol, R. L. 2007. A comprehensive tool for modeling CMOS image-sensor-noise performance. *IEEE Transactions on Electron Devices (T-ED)*, 54(6): 1321–1329.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1730–1739.

Janesick, J.; Klaasen, K.; and Elliott, T. 1985. CCD charge collection efficiency and the photon transfer technique. *The International Society for Optical Engineering (SPIE)*, 570: 7–19.

Joiner, B. L.; and Rosenblatt, J. R. 1971. Some properties of the range in samples from Tukey's symmetric lambda distributions. *American Statistical Association (ASA)*, 66(334): 394–399.

Konnik, M.; and Welsh, J. 2014. High-level numerical simulations of noise in CCD and CMOS photosensors: review and tutorial. *arXiv preprint arXiv:1412.4031*.

Low, W. F.; and Lee, G. H. 2023. Robust e-NeRF: NeRF from sparse & noisy events under non-uniform motion. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 18335–18346.

Nam, S.; Hwang, Y.; Matsushita, Y.; and Joo Kim, S. 2016. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1683–1691.

Perazzi, F.; Pont-Tuset, J.; McWilliams, B.; Van Gool, L.; Gross, M.; and Sorkine-Hornung, A. 2016. A benchmark dataset and evaluation methodology for video object segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 724–732.

Plotz, T.; and Roth, S. 2017. Benchmarking denoising algorithms with real photographs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1586–1595.

Qian, G.; Wang, Y.; Gu, J.; Dong, C.; Heidrich, W.; Ghanem, B.; and Ren, J. S. 2022. Rethinking learning-based demosaicing, denoising, and super-resolution pipeline. In *IEEE International Conference on Computational Photography (ICCP)*, 1–12.

Wang, J.; Yu, Y.; Wu, S.; Lei, C.; and Xu, K. 2021. Rethinking Noise Modeling in Extreme Low-Light Environments. In *IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Wang, Y.; Huang, H.; Xu, Q.; Liu, J.; Liu, Y.; and Wang, J. 2020. Practical deep raw image denoising on mobile devices. In *European Conference on Computer Vision (ECCV)*, 1–16.

Wei, K.; Fu, Y.; Zheng, Y.; and Yang, J. 2021. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 44(11): 8520–8537.

Zhang, F.; Xu, B.; Li, Z.; Liu, X.; Lu, Q.; Gao, C.; and Sang, N. 2023. Towards General Low-Light Raw Noise Synthesis and Modeling. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 10820–10830.

Zhang, T.; Fu, Y.; and Li, C. 2022. Deep spatial adaptive network for real image demosaicing. In *AAAI Conference on Artificial Intelligence (AAAI)*, 3326–3334.

Zhang, Y.; Qin, H.; Wang, X.; and Li, H. 2021. Rethinking noise synthesis and modeling in raw denoising. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 4593–4601.

Zhao, J.; Xiong, R.; Liu, H.; Zhang, J.; and Huang, T. 2021. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11996–12005.

Zhao, J.; Zhang, S.; Ma, L.; Yu, Z.; and Huang, T. 2022. SpikingSIM: A bio-inspired spiking simulator. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, 3003–3007.

Zhao, R.; Xiong, R.; Zhang, J.; Yu, Z.; Zhu, S.; Ma, L.; and Huang, T. 2024a. Spike Camera Image Reconstruction Using Deep Spiking Neural Networks. *IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT)*, 34(6): 5207–5212.

Zhao, R.; Xiong, R.; Zhao, J.; Zhang, J.; Fan, X.; Yu, Z.; and Huang, T. 2024b. Boosting Spike Camera Image Reconstruction from a Perspective of Dealing with Spike Fluctuations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 24955–24965.

Zhu, L.; Dong, S.; Huang, T.; and Tian, Y. 2019. A retina-inspired sampling method for visual texture reconstruction. In *IEEE International Conference on Multimedia and Expo (ICME)*, 1432–1437.

Zhu, L.; Li, J.; Wang, X.; Huang, T.; and Tian, Y. 2021. NeuSpike-Net: High Speed Video Reconstruction via Bio-inspired Neuromorphic Cameras. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2380–2389.

Zhu, L.; Zheng, Y.; Geng, M.; Wang, L.; and Huang, H. 2023. Recurrent spike-based image restoration under general illumination. In *ACM International Conference on Multimedia (ACM MM)*, 8251–8260.