Intensity-Robust Autofocus for Spike Camera

Changqing Su[1,†] Zhiyuan Ye[2] Yongsheng Xiao[2,†] You Zhou[3] Zhen Cheng[4] Bo Xiong[1,*] Zhaofei Yu[1]
Tiejun Huang[1]
Peking University[1] Nanchang Hangkong University[2] Nanjing University[3] Tsinghua University[4]
Project page: https://github.com/Onetism/saf-code

Abstract

Spike cameras, a novel neuromorphic visual sensor, can capture full-time spatial information through spike stream, offering ultra-high temporal resolution and an extensive dynamic range. Autofocus control (AC) plays a pivotal role in a camera to efficiently capture information in challenging real-world scenarios. Nevertheless, due to disparities in data modality and information characteristics compared to frame stream and event stream, the current lack of efficient AC methods has made it challenging for spike cameras to adapt to intricate real-world conditions. To address this challenge, we introduce a spike-based autofocus framework that includes a spike-specific focus measure called spike dispersion (SD), which effectively mitigates the influence of variations in scene light intensity during the focusing process by leveraging the spike camera's ability to record full-time spatial light intensity. Additionally, the framework integrates a fast search strategy called spike-based golden fast search (SGFS), allowing rapid focal positioning without the need for a complete focus range traversal. To validate the performance of our method, we have collected a spike-based autofocus dataset (SAD) containing synthetic data and real-world data under varying scene brightness and motion scenarios. Experimental results on these datasets demonstrate that our method offers state-of-the-art accuracy and efficiency. Furthermore, experiments with data captured under varying scene brightness levels illustrate the robustness of our method to changes in light intensity during the focusing process.

1. Introduction

The spike camera is an innovative neuro-inspired visual sensor developed based on the sampling mechanism of the fovea centralis region in primate retinas [1], providing exceptionally high temporal resolution and a broad dynamic range. This makes it well-suitable for various tasks in high-speed imaging [2] and computational vision [1], such as optical flow estimation [3], depth estimation [4], super-resolution [5], and object tracking [6]. The success of these tasks relies heavily on the camera's capability to capture reliable information, a factor significantly influenced by the camera's focusing performance. Generally, images obtained at the focused position contain clearer texture information compared to those at defocused positions (Fig. 1d). A robust autofocus control (AC) is crucial for enabling the spike camera to achieve reliable perception in high-speed motion scenarios.

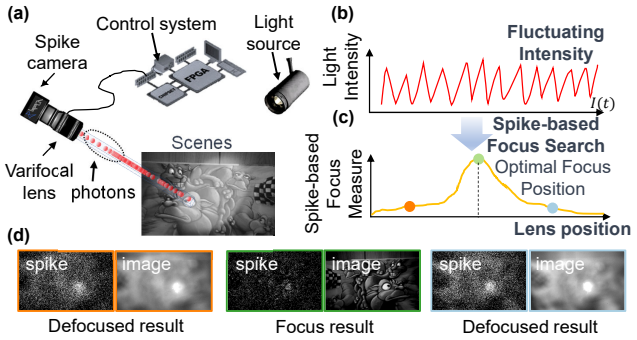

Figure 1. (a) A spike camera is equipped with a varifocal lens controlled by the control system, and can continuously record variations in scene illumination, emitting spike at a specific frequency. (b) The temporal changes in scene illumination, which determine the spike emission frequency of the spike camera, typically exhibit fluctuations. (c) The proposed spike-based measure and spike-based focus search methods can still perform well in scenarios with fluctuations in scene illumination. (d) The intensity of light at the defocused position is higher compared to the focused position, and the focused spike image is sharper and more informative than the defocused one.

*X.Bo is the corresponding author. † Contributed equally. This work was supported by the National Natural Science Foundation of China (Grant No.62371006, 62261040, 62176003, 62088102, 62071219), the China Postdoctoral Science Foundation (Grant No.2023TQ0006, GZC20230057), Beijing Nova Program (Grant No.20230484362) and Beijing Natural Science Foundation (Grant No.3242008).
[1]qing1286765276@gmail.com, {xiongbo, yuzf12, tjhuang}@pku.edu.cn, [2]{2204085400063@stu., xysfly@}nchu.edu.cn, [3]zhouyou@nju.edu.cn, [4]zcheng@mail.tsinghua.edu.cn

Traditional focus methods for frame-based cameras mainly rely on specific features of frame images, roughly divided into two categories: Spatial features [7–11] and frequency domain features [12–20]. As for event cameras, focusing methods are primarily categorized into frame-based (extended to reconstruction images from events [21–23]) and event-based [24] approaches. However, due to the differences in data modality and information characteristics compared to frame and event data, existing focusing methods cannot be directly applied to spike cameras. Moreover, these methods usually assume that scene light intensity is stable, posing challenges for their application in complex real-world scenarios. To develop an autofocus method suitable for spike cameras while being resistant to light intensity disturbances, several factors need to be considered:(1)**Focus measure function**. Existing focus measure functions, whether based on frames or event streams, can only be applied to their respective data modalities and cannot be directly applied to spike streams. (2)**Data modality**. Distinguished from image data and event streams, spike streams offer high temporal resolution while retaining rich spatial texture information. (3)**Noise**. Similar to event streams, spike streams also include noise information that is difficult to filter out. (4)**Data sizes**. Due to the rich spatio-temporal information contained in spike streams, they have higher data throughput compared to event streams, making them more challenging in terms of real-time focusing. (5)**Full-time perception of light intensity**. Given that the spike streams provide continuous recording of light intensity in space, leveraging this feature holds the promise of addressing significant challenges introduced by variations in the scene illumination during the focusing period for focus measure.

The intuitive approach is to reconstruct spike streams into images [25–29], enabling direct application to traditional frame-based autofocus methods. However, some noise and additional reconstruction time can undermine the accuracy and real-time requirements of the focusing process. On the other hand, the reconstruction process still poses similar challenges in real-world scenarios with fluctuations in light intensity (Fig. 1b). Additionally, learning-based reconstruction is often time-consuming, computing resources consuming and challenging to meet the high efficiency required for autofocus. To address these challenges, we have developed a spike-based autofocus framework. Firstly, we have developed a first spike-based focus measure leveraging the statistics of spike dispersion (SD). This measure effectively scores the spike stream at different focal positions, and mitigates the impact of changing illumination during the focusing process (Fig. 1c). Secondly, we propose a spike-based golden fast search (SGFS) approach, which, in combination with our focus measure and the cycle of the AC system (Fig. 1a), accurately positions the optimal focus without the need to traverse all focal positions. SGFS exhibits consistent parameters across different search intervals, enabling robust and rapid focusing in complex scenarios, such as fluctuations or continuous changes in light intensity. In summary, our contributions can be summarized in three main aspects:

- We propose a novel spike-based focus measure, referred to as spike dispersion (SD), for scoring the spike stream at different focal positions. This measure is characterized by its efficiency, ease of implementation, and robustness against variations in light intensity and noise.
- Integrating our spike-based focus measure and leveraging the real-time feedback from the control system, we propose a robust and efficient method to accurately locate the optimal focal position for spike cameras without the need to traverse all focal positions. Moreover, it may even be possible to skip the position initialization step in each focusing process.
- We have collected a dataset for spike-based autofocus (SAD), which includes synthetic and real data across various scenarios involving variations in scene brightness and motion. We performed comprehensive evaluations and comparisons on the SAD dataset.

2. Related Work

Focus measure. The existing focus measures can be broadly categorized into two types: **(i) Frame-based focus measure**. Frame-based focus measures can be roughly classified into four categories: gradient-based methods [7, 9, 10, 30], statistical-based methods [8, 9, 31–33], correlation-based methods [34], and transform-based methods [12–20]. Additionally, there is a recent trend of introducing methods based on deep learning [35, 36]. However, all these methods are designed for image-specific data modality. In contrast, spike streams, being binary array stream outputs rather than images, pose a challenge for the direct application of the frame-based focus measures. **(ii)Event-based focus measure**. There are two main focus measures for event cameras. The most straightforward approach is to apply frame-based focus measures directly to the reconstructed images from events [26–28]. However, these methods are either susceptible to noise or challenging to achieve real-time reconstruction on devices with limited computing resources, thus restricting their universal applications in all scenarios. To address these issues, Lin *et al*. proposed an event-based focus measure [24] from another perspective, called event rate (ER). It directly leverages the characteristics of events themselves, demonstrating excellent computational efficiency and noise robustness. However, it is a challenge for event stream in applications involving scenes with intensity changes due to the lack of complete recorded inten-

sity information. Therefore, there is still a significant demand for focus measures that leverage the intrinsic features of spike-based data and can effectively address challenges posed by variations in light intensity.

Search Methods. The rapid search method combines the results of the focus measure to achieve the optimal focal position. Traditional frame-based search methods typically entail systematically traversing various focal positions, capturing images at each location to generate a sequence of images reflecting variations in focus. Subsequently, these images are scored based on a focus measure, and search algorithms [37–41] are then applied to identify the position associated with the highest score, which indicates the optimal focal position. However, these methods work with image sequences that contain rich information, with each pixel containing extensive grayscale information. In contrast, neuromorphic cameras typically capture sparse and discrete information. Based on the characteristics of event data, Lin *et al*. proposed the event-based golden search (EGS) algorithm [24], which significantly mitigates the impact of noise on search results, demonstrating higher search efficiency and accuracy. Nevertheless, their autofocus process still involves scanning through all focal positions, resulting in data redundancy and prolonged autofocus time. Spike streams also come with a significant amount of noise. Hence, a search strategy similar to EGS can be employed, but there is still room for optimization based on the specific characteristics of the spike streams.

Spike-based image reconstruction. A naive autofocus method for spike streams involves applying frame-based focus measures directly to the reconstructed images from the spike streams. Several recent works have provided support for this method. Zhu *et al*. [26] proposed directly recovering light intensity from the statistical features of spike streams, introducing two typical reconstruction methods, namely, texture from playback (TFP) and texture from ISI (TFI). Zheng *et al*. [25] and Zhu *et al*. [28] also introduced reconstruction algorithms inspired by biological principles, leading to an enhancement in the quality of reconstruction. However, these methods often encounter a trade-off between noise and motion blur. End-to-end convolutional neural networks [29] have achieved remarkable reconstruction quality, but they depend on extensive labelled datasets, often involving intricate and time-consuming steps. Moreover, there is a potential challenge related to generalization, as the method may not adequately cover complex real-world scenarios. Chen *et al*. [42] further developed a self-supervised method to achieve high-quality reconstruction. However, neural networks still require significant computational resources, making it challenging for applications with limited computing power, such as mobile devices. Hence, the development of efficient autofocus methods based on the intrinsic characteristics of spike streams remains paramount.

3. Preliminaries

3.1. Problem Formulation

Cameras typically comprise a lens system, where parallel light converges to a point known as the focal point. The distance from the focal point to the centre of the lens is termed the focal length, denoted as f. The distance from the lens centre to the imaging object is referred to as the object distance u, while the distance from the lens centre to the imaging sensor is called the image distance v. Lens systems are often simplified as thin lenses, and the imaging model adheres to the following equation:

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f} \tag{1}$$

The focusing process essentially involves adjusting the focal position $\Gamma(t)$ by moving the lens, simultaneously altering the relationship between the object distance and image distance until it satisfies Eq. (1). At this point, the corresponding position is considered the optimal focal position $\Gamma(t^*)$. When $\Gamma(t)$ is not equal to v, defocusing occurs, and this can be typically quantified by the value of $\sigma = \Gamma(t) - v$. In the specific focusing process, the autofocus system shifts the lens along a designated search path, concurrently gathering time-synchronized data as feedback to minimize σ. Spike-based autofocus (SAF) utilizes spike streams with rich spatial and temporal information, bringing the possibility of mitigating the effects of light intensity fluctuations. SAF can estimate the optimal focal position by solving the equation:

$$\Gamma(t^*) = \underset{t}{\operatorname{argmax}} \Psi_s\left(Q_s(t,\Delta t), t\right) \tag{2}$$

where Ψ_s represents the spike-based focus measure function, Q_s represents the spike-based intensity sequence sampled during the focusing movement. In contrast to the light intensity difference sampling mode of event cameras, spike cameras are developed based on an integration sampling approach. The sampled results in Q_s consist of a set of spike image sequences $\left\{s_t : t \in \left[t-\frac{\Delta t}{2}, t+\frac{\Delta t}{2}\right]\right\}$, mainly controlled by sampling time and intervals. Each spike frame has a size of $w \times h$, where w represents the width of the sensor's pixel and h represents the height of the sensor's pixel.

3.2. Spike Formation Model

The sensor of the spike camera comprises an array of $H \times W$ pixels, where each pixel asynchronously integrates photons in the spatial domain. The photodiode then converts the photocurrent $I_{\boldsymbol{x}}$ into the voltage $V_{\boldsymbol{x}}$ on capacitance C_p. Once the voltage $V_{\boldsymbol{x}}$ reaches a predefined threshold ϑ, it

triggers the output of a spike while simultaneously resetting the photodiode. The entire process can be formulated as follows:

$$\frac{1}{C_p}\int_{t}^{t+\Delta t} I_{\boldsymbol{x}} dt \geq \vartheta \tag{3}$$

where Δt represents the integration period time. It can be derived that the firing frequency of spikes is proportional to the light intensity, where higher brightness results in a higher frequency of spikes.

4. Methods

4.1. Power measure

The fundamental representation of power measure pertains to describing the power of electrical circuit components. Taking a resistor R as an example, when voltage $U(t)$ passes through the resistor, the instantaneous power is defined as $U(t)^2/R$ and the overall average power is defined as follows:

$$P = \frac{1}{T}\int_{0}^{T} U(t) * A(t) dt \tag{4}$$

where $A(t) = \frac{U(t)}{R}$. The power can primarily be divided into the direct current (DC) component and the alternating current (AC) component [7]. The AC component primarily considers the power with the temporal variation of $U(t)$, reflecting the stability of the system. Inspired by this, a similar image power measure based on the AC component has been proposed in traditional frame-based focusing approaches [7], which can be defined as:

$$M_i = \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}(G(\boldsymbol{x}) - u)^2 \tag{5}$$

where M_i is frame-based focus measure, $G(\boldsymbol{x})$ is the pixel value at position $\boldsymbol{x} = (x, y)$ in the image, N is the number of pixels in the image and u is the mean intensity of the image. During the focusing process, the power of the DC component is generally considered to remain relatively constant, while the AC component tends to increase as the focus gradually improves. By leveraging this characteristic, frame-based focusing methods can employ optimization to locate the optimal focal position. However, for the spike camera, it remains an open challenge to find an efficient spike-based measure to measure spike streams.

4.2. Spike-based Focus Measure: Spike dispersion

The spike-based focus measure primarily aims to efficiently quantify and score spike streams captured at different focal positions. However, traditional methods like gradient-based measures or event-based measures such as event rate, can only be applied to specific data modalities and are not directly applicable to spike data. Although recent learning-based approaches can effectively reconstruct images from spike streams, these processes are time-consuming and demand substantial computational resources. As a solution to this difficulty, Zhu *et al*. [26] proposed TFP method approximates intensity by directly accumulating spike streams, enabling rapid reconstruction. However, as discussed in [42], it faces challenges in balancing between noise and motion blur, making it difficult to apply to challenging real-world scenarios.

To address these challenges, we propose employing spike dispersion (SD) as a spike-based focus measure. Inspired by the concept of AC power in circuits, spike dispersion measures the AC component in the overall power on the sensor. SD only requires straightforward calculations on the spike streams without need for additional reconstruction into image-type data. Due to the continuous spatial-temporal recording capability of the spike stream, it theoretically can recover the light intensity information at any given moment. By leveraging this feature, SD can effectively withstand challenges posed by variations in light intensity in the scene. Our proposed SD primarily calculates the normalized AC component of the overall power on the sensor within a time interval Δt, expressed as:

$$R_s(t, \Delta t) = \frac{1}{u^2}\sum_{\boldsymbol{x}\in\Omega}\left(\int_{t-\Delta t/2}^{t+\Delta t/2} S_{\boldsymbol{x}}(t)dt - u\right)^2 \tag{6}$$

where $u = \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}\int_{t-\Delta t/2}^{t+\Delta t/2} S_{\boldsymbol{x}}(t)dt$ is the average power within the time interval Δt. $S_{\boldsymbol{x}}(t)$ is the spike value at position $\boldsymbol{x} = (x, y)$ at the time t. We will provide a detailed explanation of why SD is an effective spike-based focus measure. In Eq. (3), it is generally assumed that the light intensity remains constant within the small interval Δt, thereby making $I_{\boldsymbol{x}}$ approximately be a constant. We can obtain the approximate voltage for a pixel at $\boldsymbol{x} = (x, y)$ by:

$$U_{\boldsymbol{x},t} = \frac{1}{C_p}\left(I_{\boldsymbol{x},t} \times \Delta t\right) \approx \vartheta\int_{t}^{t+\Delta t} S_{\boldsymbol{x}}(t)dt \tag{7}$$

where $S_{\boldsymbol{x}}(t)$ is the spike value at position $\boldsymbol{x} = (x, y)$ at the time t, $I_{\boldsymbol{x},t}$ is the photocurrent in the interval Δt at the time t. Based on Eq. (4), we can further derive the average power for a pixel as:

$$P_{\boldsymbol{x},t} = U_{\boldsymbol{x},t} * I_{\boldsymbol{x},t} = I_{\boldsymbol{x},t} \times \vartheta\int_{t}^{t+\Delta t} S_{\boldsymbol{x}}(t)dt \tag{8}$$

By changing the interval from $[t, t+\Delta t]$ to $[t-\Delta t/2, t+\Delta t/2]$, the average power $P_{u,t}$ on the sensor at time t can be expressed as:

$$P_{u,t} = \frac{I_{\boldsymbol{x},t} \times \vartheta}{N}\sum_{\boldsymbol{x}\in\Omega}\int_{t-\Delta t/2}^{t+\Delta t/2} S_{\boldsymbol{x}}(t)dt \tag{9}$$

where N is the number of pixels on the sensor. According to the Eq. (5), a similar spike-based power measure M_s can be derived:

$$M_s = \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}\left(P_{\boldsymbol{x},t} - P_{u,t}\right)^2$$
$$= \frac{\left(I_{\boldsymbol{x},t}\times\vartheta\right)^2}{N}\sum_{\boldsymbol{x}\in\Omega}\left(S_{\boldsymbol{x},\Delta t}(t) - \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}S_{\boldsymbol{x},\Delta t}(t)\right)^2 \tag{10}$$

where $S_{\boldsymbol{x},\Delta t}(t) = \int_{t-\Delta t/2}^{t+\Delta t/2} S_{\boldsymbol{x}}(t)dt$. however, $I_{\boldsymbol{x},t}$ is correlated with the light intensity of the scene. Its presence leads to the disturbance of the AC component by the light intensity, making it challenging to cope with actual fluctuations in the scene's light intensity. Therefore, we divided Eq. (10) by the power of Eq. (9), thereby eliminating the influence of light intensity, expressed as:

$$\frac{M_S}{{P_{u,t}}^2} = \frac{\frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}\left(s_{\boldsymbol{x},\Delta t}(t) - \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}S_{\boldsymbol{x},\Delta t}(t)\right)^2}{\left(\frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}S_{\boldsymbol{x},\Delta t}(t)\right)^2} \tag{11}$$

Letting $u = \frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}S_{\boldsymbol{x},\Delta t}(t)$, Eq. (11) can be further expressed as:

$$\frac{M_s}{{P_{u,t}}^2} = \frac{\frac{1}{N}\sum_{\boldsymbol{x}\in\Omega}\left(\int_{t-\Delta t/2}^{t+\Delta t/2}S_{\boldsymbol{x}}(t)dt - u\right)^2}{u^2} \propto R_s(t,\Delta t) \tag{12}$$

According to Eq. (12), $R_s(t,\Delta t)$ is proportional to the AC component of power on the sensor. implying that SD can effectively characterize the variations in the AC power. As mentioned in Sec. 4.1, the AC component of power is typically maximal at the focal position. Due to the proportional relationship between SD and the AC component, SD also reaches its maximum at the focal position. Therefore, SD can serve as an effective spike-based focus measure, computed as:

$$\Psi_s\left(Q_s(t,\Delta t),t\right) \overset{\text{def}}{\Longrightarrow} R_s(t,\Delta t) \tag{13}$$

4.3. Optimization

Given the spike-based focus measure in Eq. (13), we can identify the optimal focal position by solving the problem in Eq. (2). During optimization, we mainly introduce three approaches. The first is the most primitive method that requires a manual choice of Δt to compute SD and then is an extension of the EAF method to the spike stream. The last is our proposed SGFS method, incorporating the advantages of EAF but with higher efficiency.

Naive SAF with Manually Chosen Δt. According to Eq. (6), once Δt is determined, we can calculate the focus score at any given moment t. Based on this feature, combined with real-time feedback from the focus traversal process, we can rapidly identify the moment t with the maximum focus score as the optimal t^*. The detailed algorithmic process is outlined in [24].

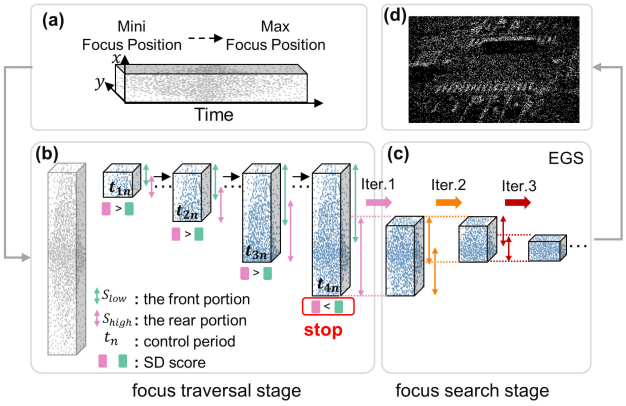

Figure 2. (a) Our spike-based autofocus system initiates without any initial data, and then, (b) employing real-time computed SD, it determines whether to continue data collection at the current position without the need to traverse all focal positions. (c) Upon meeting the stopping criteria, we employ a method similar to EGS to search within the collected data for the optimal focal position, and (d) adjust the lens accordingly.

EAF extend to SAF. To eliminate the impact of the Δt parameter chosen in EAF, Lin *et al.* [24] proposed the Event-based golden Search (EGS) algorithm, which can adaptively adjust the Δt parameter during the search process, significantly mitigating the influence of noise on the focusing process. This algorithm can also be extended to SAF. The process rapidly identifies the optimal focal position while mitigating the impact of noise. However, it involves traversing all focal positions each time and requires initializing the lens position during each focusing process, thus increasing the time consumption of the focusing process.

SAF with real-time feedback. To fully leverage the real-time feedback performance of the AC system while eliminating the impact of noise, we developed the spike-based golden fast search (SGFS) based on the event-based golden search (EGS) [24]. SGFS allows for adaptive parameter adjustment while efficiently utilizing the available time during the traversal process for computation and feedback optimization, as shown in Fig. 2.

The overall algorithm is summarized in Algorithm 1. The key difference is that SGFS incorporates real-time feedback calculations, which enables the focus system to more quickly locate the optimal focal position without traversing all possible focal positions. During the focus traversal process, the subsequent focal positions can be determined as defocused positions without further traversal, when the overall energy dissipation decreases. Furthermore, by com-

paring the change in energy dissipation at the initial position, the system could better determine the direction of focus traversal without requiring a position initialization step.

Algorithm 1 : Spike-based Golden Fast Search (SGFS)

Input: threshold u, golden ratio α, control period t_c.
Output: Optimal focal position $\Gamma(t^*)$

$T = 0$;
repeat
  $T = T + t_c$; $t_1 = \alpha T/2$; $t_2 = T - \alpha T/2$; $\Delta t = \alpha T$;
until $\Psi_s(Q_s(t_1, \Delta t), t) < \Psi_s(Q_s(t_2, \Delta t), t)$
$T_1 = T - \alpha T/2$; $T_2 = T$; $T = T_2 - T_1$;
$t^* = EGS(u, \alpha, T_1, T_2)$;

5. Experiments

5.1. The Spike-based Autofocus Dataset

Synthetic Dataset: The synthetic data primarily comes from the camel and crossroad datasets in [29], representing static and dynamic scenes, respectively. To generate synthetic data, we extracted frames capturing the focus variation process from videos and linearly interpolated them to create 20,000 frames. Subsequently, we simulated spike streams using the methods outlined in [43], adjusting the intensity relationships between input frames to generate spike streams for scenarios with varying illumination conditions.

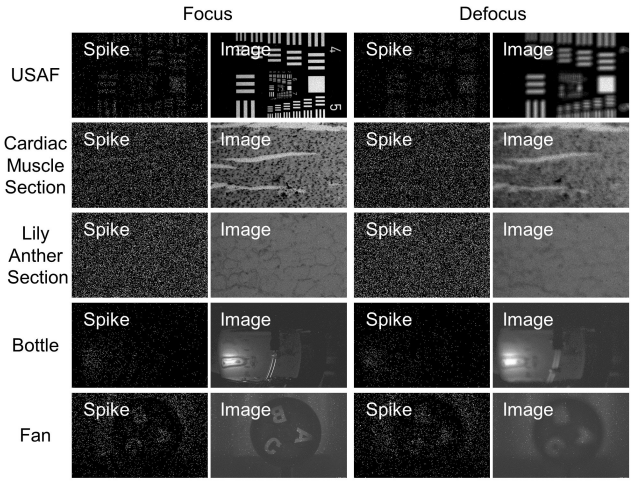

Figure 3. Examples of the spike-based autofocus dataset. Our dataset contains spike streams captured in various scenes and conditions.

Real-world Dataset: The real-world data primarily employed spike cameras mentioned in [43], mounted on an electric microscope for capturing microscopic data and directly connected to adjustable-focus lenses for capturing macroscopic data. Microscopic data was captured by electrically controlling the objective-to-sample distance, while macroscopic data was obtained by manually adjusting the adjustable focus lenses. Some samples are depicted in Fig. 3. The dynamics of the microscopic scene were achieved by controlling the electric displacement platform's movement. In the macroscopic scenario, the bottle naturally involved the movement of bubbles, while the fan was capable of rapid rotation. In the USAF scenario, datasets for three lighting conditions were captured by controlling the intensity changes of the microscope: constant, continuous variation, and fluctuations. Other samples were captured in normal environmental settings.

5.2. Quantitative Results of SD on Simulation Data

To validate the theoretical performance of the spike-based focus measure, we initially experimented with simulated data. In this experiment, the spike-based focus measure was compared with frame-based measures (including gradient-based and MF-DCT [16]), as well as the direct extension of ER [24] to the spike stream. The frame-based measures were computed based on TFP reconstruction for efficiency. The accumulated values within a certain time interval on the sensor can reflect the trend of changes in scene light intensity. In all subsequent experiments, we utilize these accumulated values to characterize the variation in scene light intensity. We simulated three common light intensity scenarios on simulated data: constant, continuous variations, and sudden fluctuations. In these three scenarios, the curves of normalized focus scores with changes in focal position for different focus measures are illustrated in Fig. 4. In static scenes ("Constant" column in Fig. 4(a)), all three methods perform well when the light intensity is constant. However, in dynamic scenes ("Constant" column in Fig. 4(b)), MF-DCT and ER fail to function properly. During the focusing process with continuous changes in light intensity, the dominant influence on the measurement results for methods other than SD is the impact of light intensity changes, rather than the effect of focusing changes. This poses a challenge for these focusing measures to operate effectively in such scenarios. Although SD may be influenced to some extent, it generally remains within a functional range, as shown in the "Continuous variations" column in Fig. 4. In scenes with light intensity fluctuations, except for SD, other methods are completely unable to function properly. Fluctuations in light intensity can completely disrupt the original trends of measurement methods with changes in focal position. SD, however, works effectively in this scenario, as demonstrated in the "Sudden changes" column in Fig. 4. We also present results under varying light intensity levels in the supplementary materials, demonstrating similar results.

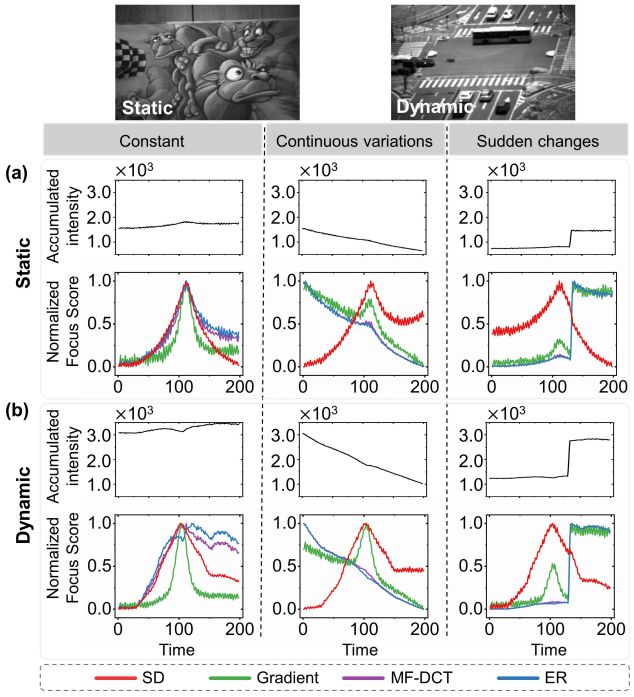

Figure 4. Focus scores in simulation data (including static (a) and dynamic (b) scenes, along with variations in light intensity) of the proposed spike-based focus measure, *e.g.*, the spike dispersion (SD), the frame-based focus measure *e.g.*, gradient-based and MF-DCT [16], and extension of ER [24] to spike stream. SD performs well in all scenarios, while other methods may face challenges in specific situations, particularly in scenes with light intensity fluctuations, where only SD remains effective.

5.3. Quantitative Results of SD on Real-world Data

To comprehensively validate the performance of the focus measure on real-world data, we captured focusing scenes under an electrically controlled microscope on the USAF sample. The dataset includes scenes that are completely consistent with the simulated data, comprising two scenes (static and dynamic) and three lighting conditions (constant, continuous variations, and sudden fluctuations). In various scenarios, the curves of normalized focus scores with changes in focal position under different focus measures are shown in Fig. 5. When the scene brightness is constant, all methods can achieve good measurement performance, as shown in the "Constant " column of Fig. 5. However, in a static scene with low light intensity ("Constant" column in Fig. 5(a)), an increase in noise and relatively less structural information in reconstructed images may lead to a decline in frame-based measurement performance, and gradient-based methods may even fail to operate effectively. In a scenario with continuous changes in scene brightness ("Continuous variations " column of Fig. 5), similar to the simulation results, MF-DCT and ER are significantly affected by changes in light intensity, rendering them ineffective. Gradient-based methods also experience some impact but still exhibit good performance. In contrast, SD consistently maintains robust performance. ("Continuous variations " column of Fig. 5(b)). In scenes with fluctuations in scene brightness, methods other than SD completely fail to function properly, and in some cases, they exhibit a reverse trend in this scenario. However, SD consistently demonstrates good performance, as shown in the "Sudden changes" column of Fig. 5. The results demonstrate that SD effectively mitigates the impact of scene lighting changes on focusing performance. Results regarding additional light intensity variations can be found in the supplementary materials, demonstrating consistent performance.

5.4. Comparison of SGFS and EGS

To further analyse the performance of our proposed method in real-world data, we compared spike-based focus measures with frame-based measures (including gradient-based

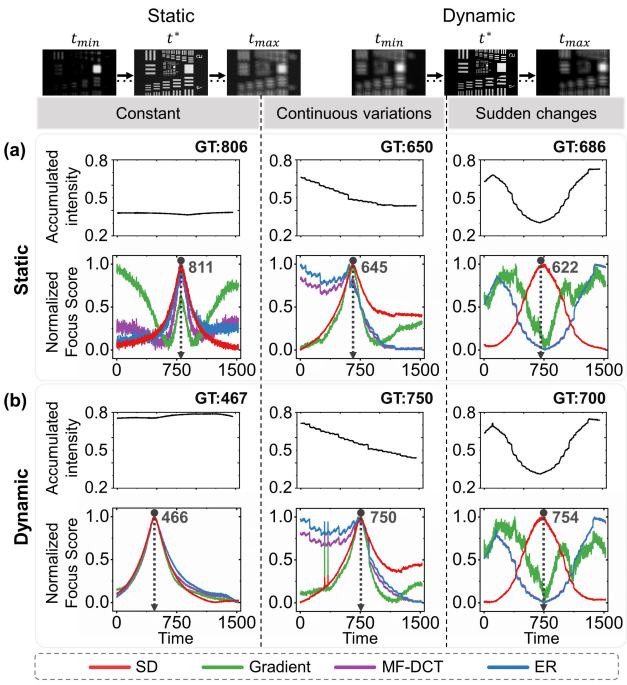

Figure 5. Focus scores in the USAF data (including static (a) and dynamic (b) scenes, along with variations in light intensity) of the proposed spike-based focus measure, *e.g.*, the spike dispersion (SD), the frame-based focus measure *e.g.*, gradient-based and MF-DCT [16], and extension of ER [24] to spike stream. The arrows with numerical markings indicate the predicted focusing position, while the ground truth values correspond to the GT values in the top right corner. SD performs well in all scenarios, while other methods may face challenges in specific situations, particularly in scenes with light intensity fluctuations, where only SD remains effective.

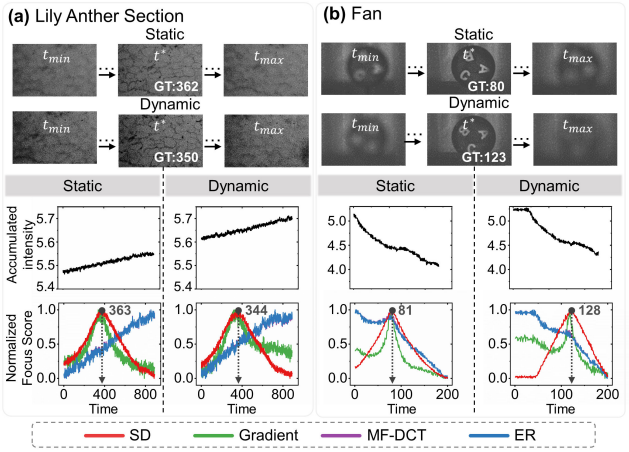

Figure 6. Focus scores in the microscopic data (Lily Anther Section (a)) and macroscopic data (Fan (b)) data of the proposed spike-based focus measure, *e.g.*, the spike dispersion (SD), the frame-based focus measure *e.g.*, gradient-based and MF-DCT [16], and extension of ER [24] to spike stream. In the final row of the figure, the arrows with numerical markings indicate the predicted focusing position, while the ground truth values are marked by the GT values in the rows above. SD performs well in all scenarios, However, since this focusing process involves continuous changes in light intensity, other methods still face challenges in these scenes.

and MF-DCT [16]) and the direct extension of ER [24] to spike streams (Fig. 6(a,b)). Additionally, we evaluated the performance of our SGFS search method in comparison to EGS [24] (Fig. 7). For the real-world data from the lily anther section and fan scenarios, the fluctuations in light intensity resembled scenes with continuous changes. In such cases, MF-DCT and ER are almost dominated by changes in light intensity and fail to function properly. Gradient-based methods also exhibited instability, being affected by variations in light intensity. In contrast, SD outperforms other methods, demonstrating superior robustness to both light intensity and noise (Fig. 6(a-b)). SD+SFGS demonstrates outstanding performance, wherein SFGS, through the integration of real-time computed SD and leveraging prior knowledge, effectively stop the traversal process at the appropriate position without the need to traverse all focal positions. The left column of Fig. 7(a-b) shows the position where the traversal could be stopped, indicated by the red dot. The stopping position is typically close to the optimal focal position. Subsequently, Continuing with a method similar to EGS, we further searched for the optimal focal position within the previously traversed locations. Its convergence speed is faster than the global traversal of EGS, typically being more than twice as fast, as shown in the right column of Fig. 7(a) and Fig. 7(b). We have also included comparison results for additional samples in the supplementary material.

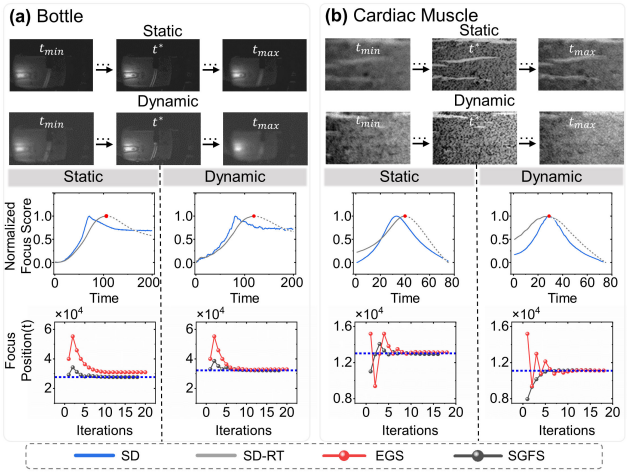

Figure 7. Performance comparison of SGFS and EGS in the macroscopic data (Bottle (a)) and microscopic data (Cardiac Muscle Section (b)). The blue dashed line in the final row of the figure represents the actual focusing position. The left column of (a) and (b) shows the focus scores of SD and the real-time feedback SD (SD-RT) during the direct traversal of focal positions. According to the SD-RT results, the traversal can be stopped at the position of the red dot, which implies that data collection is unnecessary for the grey area. Subsequently, the optimal focal position is determined using the similar EGS method. The right column of (a) and (b) illustrates the changes in focal positions during the search process. Compared to the global search of EGS, SGFS converges more quickly, typically at twice the speed of EGS.

6. Conclusion and Discussion

In this paper, we propose the first spike-based solution for automatic focusing tasks, comprising a simple and efficient focusing measure robust to changes in light intensity, referred to as spike dispersion (SD), and a spike-based golden search (SGFS) that rapidly locates the optimal focal position without the need to traverse all focal positions. We also collected a dataset containing simulated and real-world data under various lighting conditions and motion scenarios. Both the simulated and captured data confirm that our approach can achieve spike camera focusing in complex real-world scenes, particularly in scenarios with light intensity fluctuations.

Limitations. Addressing the issue of focusing on a scene with multiple targets typically requires the integration of target detection and recognition technologies. Our method is an initial exploration focused solely on automatic focusing with spike cameras, yet its performance in such complex scenarios is currently subpar. we plan to extend our method for a wider range of scene requirements in future work.

References

[1] Tiejun Huang, Yajing Zheng, Zhaofei Yu, Rui Chen, Yuan Li, Ruiqin Xiong, Lei Ma, Junwei Zhao, Siwei Dong, Lin Zhu, Jianing Li, Shanshan Jia, Yihua Fu, Boxin Shi, Si Wu, and Yonghong Tian. 1000× faster camera and machine vision with ordinary devices. *Engineering*, 25:110–119, 2023. 1

[2] Yajing Zheng, Lingxiao Zheng, Zhaofei Yu, Tiejun Huang, and Song Wang. Capture the moment: High-speed imaging with spiking cameras through short-term plasticity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8127–8142, 2023. 1

[3] Rui Zhao, Ruiqin Xiong, Jing Zhao, Zhaofei Yu, Xiaopeng Fan, and Tiejun Huang. Learning optical flow from continuous spike streams. *Advances in Neural Information Processing Systems*, 35:7905–7920, 2022. 1

[4] Jiyuan Zhang, Lulu Tang, Zhaofei Yu, Jiwen Lu, and Tiejun Huang. Spike transformer: Monocular depth estimation for spiking camera. In *European Conference on Computer Vision*, pages 34–52, 2022. 1

[5] Xijie Xiang, Lin Zhu, Jianing Li, Yixuan Wang, Tiejun Huang, and Yonghong Tian. Learning super-resolution reconstruction for high temporal resolution spike stream. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):16–29, 2023. 1

[6] Yajing Zheng, Zhaofei Yu, Song Wang, and Tiejun Huang. Spike-based motion estimation for object tracking through bio-inspired unsupervised learning. *IEEE Transactions on Image Processing*, 32:335–349, 2023. 1

[7] Lawrence Firestone, Kitty Cook, Kevin Culp, Neil Talsania, and Kendall Preston Jr. Comparison of autofocus methods for automated microscopy. *Cytometry*, 12(3):195–206, 1991. 2, 4

[8] Loïc A. Royer, William C. Lemon, Raghav K. Chhetri, Yinan Wan, Michael Coleman, Eugene W. Myers, and Philipp J. Keller. Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nature Biotechnology*, 34(12):1267–1278, 2016. 2

[9] N. Ng Kuang Chern, Poo Aun Neow, and M.H. Ang. Practical issues in pixel-based autofocusing for machine vision. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, volume 3, pages 2791–2796 vol.3, 2001. 2

[10] Jan-Mark Geusebroek, Frans Cornelissen, Arnold W.M. Smeulders, and Hugo Geerts. Robust autofocusing in microscopy. *Cytometry*, 39(1):1–9, 2000. 2

[11] S.K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831, 1994. 2

[12] Matej Kristan, Janez Perš, Matej Perše, and Stanislav Kovačič. A bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform. *Pattern Recognition Letters*, 27(13):1431–1439, 2006. 2

[13] Yibin Tian, Kevin Shieh, and Christine F. Wildsoet. Performance of focus measures in the presence of nondefocus aberrations. *JOSA A*, 24(12):B165–B173, 2007.

[14] Kanjar De and V. Masilamani. Image sharpness measure for blurred images in frequency domain. *Procedia Engineering*, 64:149–158, 2013.

[15] Jaroslav Kautsky, Jan Flusser, Barbara Zitová, and Stanislava Šimberová. A new wavelet-based measure of image focus. *Pattern Recognition Letters*, 23(14):1785–1794, 2002.

[16] Sang-Yong Lee, Yogendera Kumar, Ji-Man Cho, Sang-Won Lee, and Soo-Won Kim. Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(9):1237–1246, 2008. 6, 7, 8

[17] Sang-Yong Lee, Jae-Tack Yoo, Yogendera Kumar, and Soo-Won Kim. Reduced energy-ratio measure for robust autofocusing in digital camera. *IEEE Signal Processing Letters*, 16(2):133–136, 2009.

[18] Ran Tao, Wei Zhang, and Yanlei Li. Time–frequency filtering-based autofocus. *Signal Processing*, 91(6):1401–1408, 2011.

[19] Hui Xie, Weibin Rong, and Lining Sun. Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 229–234, 2006.

[20] Ge Yang and B.J. Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, volume 3, pages 2143–2148 vol.3, 2003. 2

[21] Christian Brandli, Lorenz Muller, and Tobi Delbruck. Real-time, high-speed video decompression using a frame- and event-based DAVIS sensor. In *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 686–689, 2014. 2

[22] Gottfried Munda, Christian Reinbacher, and Thomas Pock. Real-time intensity-image reconstruction for event cameras using manifold regularisation. *International Journal of Computer Vision*, 126(12):1381–1393, 2018.

[23] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(06):1964–1980, 2021. 2

[24] Shijie Lin, Yinqiang Zhang, Lei Yu, Bin Zhou, Xiaowei Luo, and Jia Pan. Autofocus for event cameras. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16323–16332, 2022. 2, 3, 5, 6, 7, 8

[25] 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 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6354–6363, 2021. 2, 3

[26] Lin Zhu, Siwei Dong, Tiejun Huang, and Yonghong Tian. A retina-inspired sampling method for visual texture reconstruction. In *2019 IEEE Internation Conference on Multimedia and Expo (ICME)*, pages 1432–1437, 2019. 2, 3, 4

[27] Lin Zhu, Siwei Dong, Jianing Li, Tiejun Huang, and Yonghong Tian. Retina-like visual image reconstruction via

spiking neural model. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1435–1443, 2020.

[28] Lin Zhu, Jianing Li, Xiao Wang, Tiejun Huang, and Yonghong Tian. Neuspike-net: High speed video reconstruction via bio-inspired neuromorphic cameras. In *2021 IEEE/CVF Internationa Conference on Computer Vision (ICCV)*, pages 2380–2389, 2021. 2, 3

[29] Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. Spk2imgnet: Learning to reconstruct dynamic scene from continuous spike stream. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11991–12000, 2021. 2, 3, 6

[30] J F Brenner, B S Dew, J B Horton, T King, P W Neurath, and W D Selles. An automated microscope for cytologic research a preliminary evaluation. *Journal of Histochemistry & Cytochemistry*, 24(1):100–111, 1976. 2

[31] Shuxin Liu, Manhua Liu, and Zhongyuan Yang. An image auto-focusing algorithm for industrial image measurement. *EURASIP Journal on Advances in Signal Processing*, 2016(1):70, 2016. 2

[32] Chong-Yaw Wee and Raveendran Paramesran. Measure of image sharpness using eigenvalues. *Information Sciences*, 177(12):2533–2552, 2007.

[33] P. T. Yap and P. Raveendran. Image focus measure based on chebyshev moments. *IEE Proceedings - Vision, Image and Signal Processing*, 151(2):128–136, 2004. 2

[34] D. Vollath. The influence of the scene parameters and of noise on the behaviour of automatic focusing algorithms. *Journal of Microscopy*, 151(2):133–146, 1988. 2

[35] Chengyu Wang, Qian Huang, Ming Cheng, Zhan Ma, and David J. Brady. Deep learning for camera autofocus. *IEEE Transactions on Computational Imaging*, 7:258–271, 2021. 2

[36] C. Herrmann, R. Strong Bowen, N. Wadhwa, R. Garg, Q. He, J. T. Barron, and R. Zabih. Learning to autofocus. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2227–2236, 2020. 2

[37] Jie He, Rongzhen Zhou, and Zhiliang Hong. Modified fast climbing search auto-focus algorithm with adaptive step size searching technique for digital camera. *IEEE Transactions on Consumer Electronics*, 49(2):257–262, 2003. 3

[38] N. Kehtarnavaz and H. J. Oh. Development and real-time implementation of a rule-based auto-focus algorithm. *Real-Time Imaging*, 9(3):197–203, 2003.

[39] Eric P. Krotkov. *Active Computer Vision by Cooperative Focus and Stereo*. Springer Science & Business Media, 2012.

[40] Chengyu Wang, Qian Huang, Ming Cheng, Zhan Ma, and David J. Brady. Intelligent autofocus, 2020. arXiv:2002.12389 [eess].

[41] Y. Xiong and S.A. Shafer. Depth from focusing and defocusing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 68–73, 1993. 3

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

[43] Junwei Zhao, Shiliang Zhang, Lei Ma, Zhaofei Yu, and Tiejun Huang. Spikingsim: A bio-inspired spiking simulator. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 3003–3007, 2022. 6