Intensity-Robust Autofocus for Spike Camera

Supplementary Material

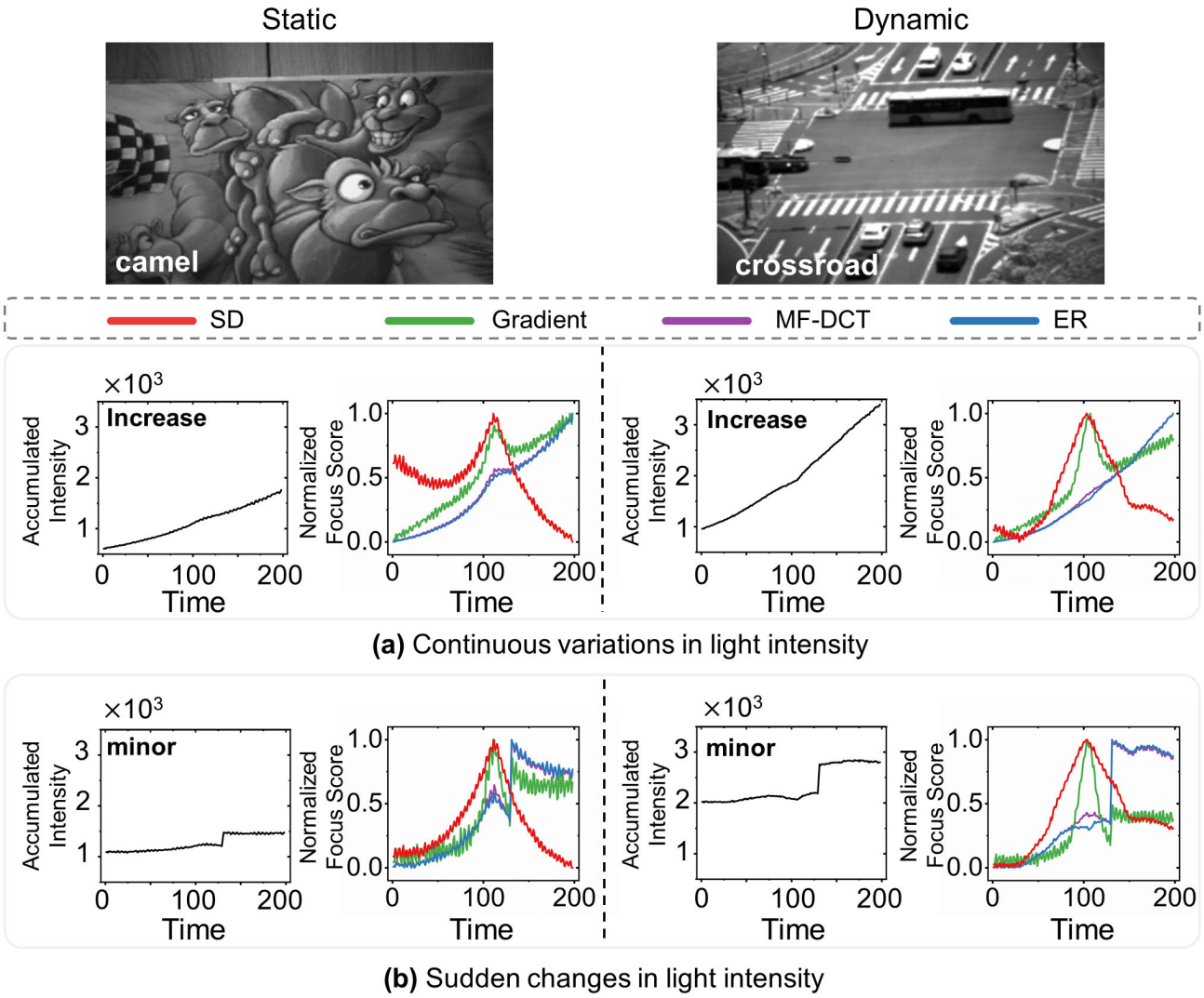

Figure 1. Focus scores in simulation data (including static and dynamic scenes, along with variations in light intensity: (a) continuous variations and (b) sudden fluctuations.) of the proposed spike-based focus measure, *e.g.* the spike dispersion (SD), the frame-based focus measure, *e.g.* gradient-based and MF-DCT [1], and extension of ER [2] to spike stream. (a) 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 focus measures to operate effectively in such scenarios. While SD may also be influenced to some extent, it generally remains within a functional range. (b) In scenes with light intensity fluctuations, except for SD, other methods are completely unable to function properly. Even minor 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.

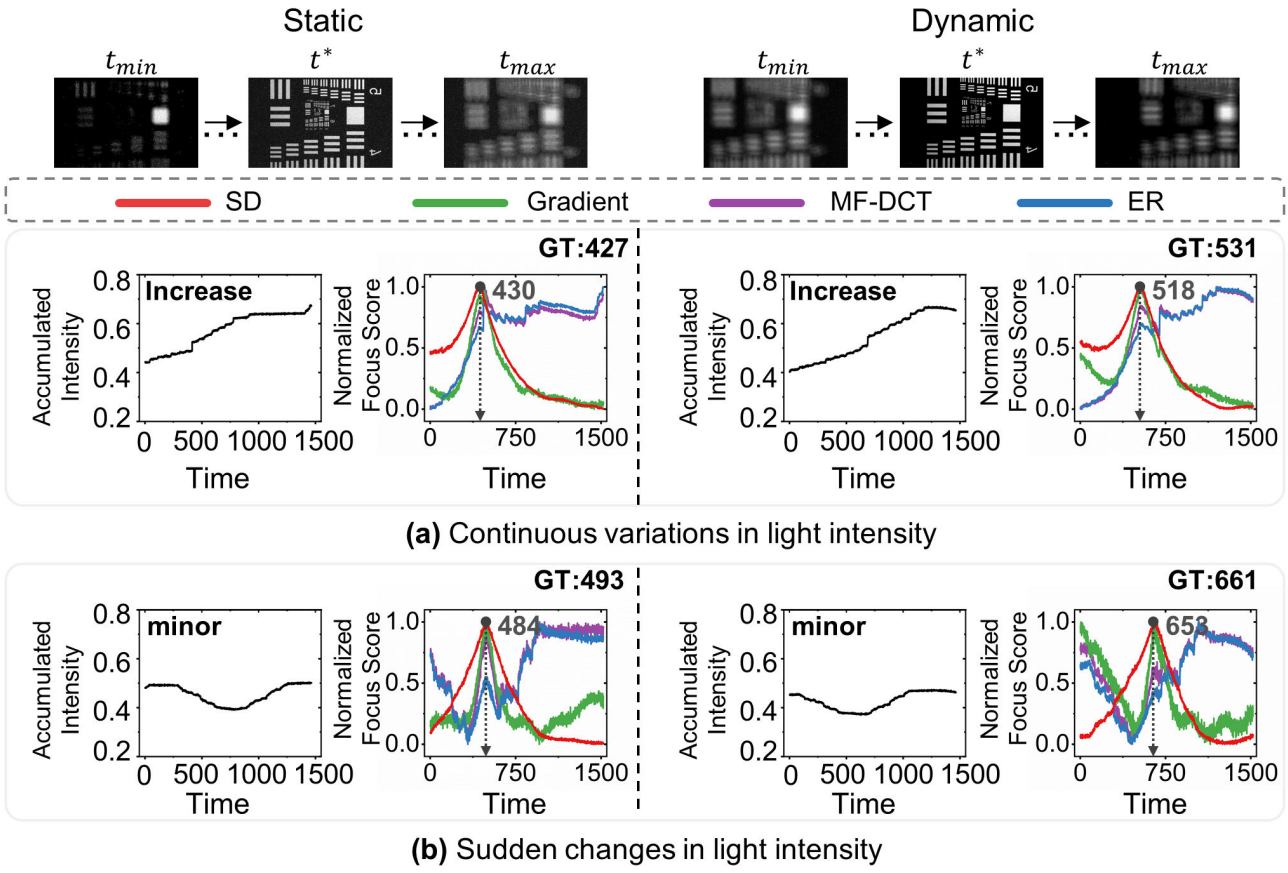

Figure 2. Focus scores in the USAF data (including static and dynamic scenes, along with variations in light intensity: (a) continuous variations and (b) sudden fluctuations.) of the proposed spike-based focus measure, *e.g.* the spike dispersion (SD), the frame-based focus measure, *e.g.* gradient-based and MF-DCT [1], and extension of ER [2] 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. (a) In scenes with continuous changes in brightness, similar to the simulation results, except for SD, other methods are affected by the light intensity, with a more pronounced impact in dynamic scenarios. (b) In a scenario with fluctuations in scene brightness, even subtle fluctuations affect methods other than SD, leading to abnormal measurements. SD, on the other hand, consistently demonstrates good performance in such scenarios.

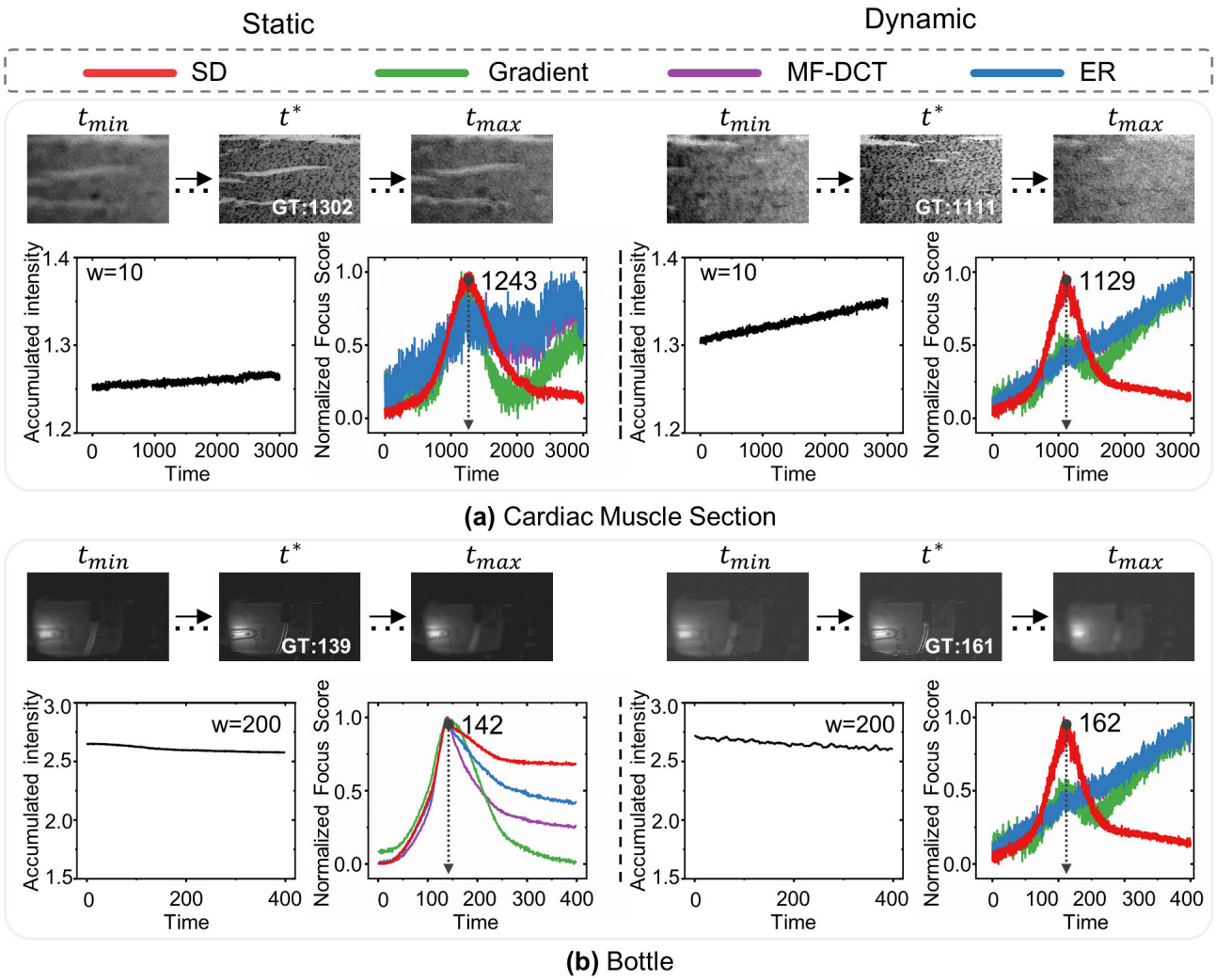

Figure 3. Focus scores in the microscopic data (Cardiac Muscle Section (a)) and macroscopic data (bottle (b)) data of the proposed spike-based focus measure, *e.g.* the spike dispersion (SD), the frame-based focus measure, *e.g.* gradient-based and MF-DCT [1], and extension of ER [2] to spike stream. 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 these focusing processes involve continuous changes in light intensity, other methods still face challenges in these scenes.

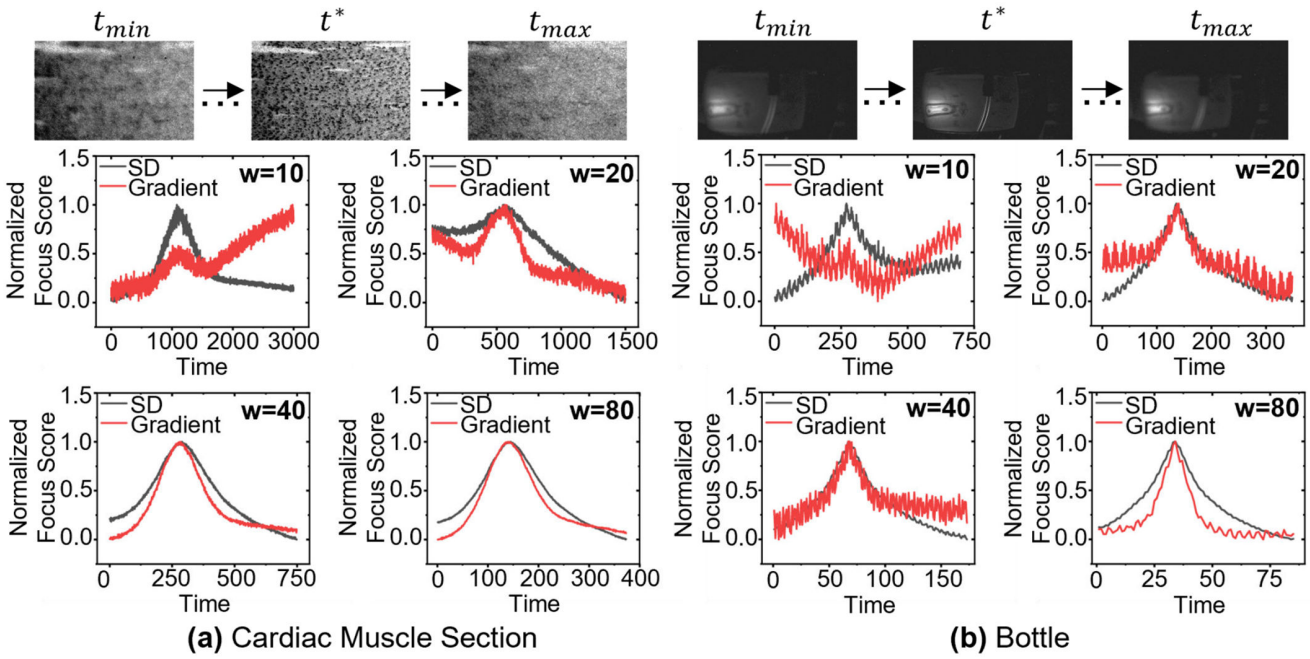

Figure 4. Focus scores in the microscopic data (Cardiac Muscle Section (a)) and macroscopic data (bottle (b)) data of the proposed spike-based focus measure, *e.g*. the spike dispersion (SD), the frame-based focus measure, *e.g*. gradient-based, where w is the size of the integration window. With a smaller window size, indicating a reduced temporal resolution, gradient-based methods exhibit inferior measurement performance. As the window size gradually increases, the accumulated information within the window approaches that of the entire frame image, resulting in an enhancement of the measurement performance for gradient-based methods. However, SD consistently demonstrates robust measurement performance under any window size. In contrast to gradient-based methods, SD operates at higher temporal resolutions, making it well-suited for high-speed scenarios. The numbers and arrow markers indicate the predicted focusing position, while the ground truth values are marked by the GT values in the rows above.

ID	Scene	Spike frames	Focal frame	Light	microscopic/macroscopic	Motion	Name
1	USAF	45600	24190	Constant	microscopic		USAF_static_constant
2	USAF	45600	19511	Decrease	microscopic		USAF_static_decrease
3	USAF	45600	12830	Increase	microscopic		USAF_static_increase
4	USAF	45600	14800	Fluctuation_minor	microscopic		USAF_static_minor
5	USAF	45600	20580	Fluctuation_significant	microscopic		USAF_static_significant
6	USAF	45600	14000	Constant	microscopic	✓	USAF_dynamic_constant
7	USAF	45600	22510	Decrease	microscopic	✓	USAF_dynamic_decrease
8	USAF	45600	15950	Increase	microscopic	✓	USAF_dynamic_increase
9	USAF	45600	19833	Fluctuation_minor	microscopic	✓	USAF_dynamic_minor
10	USAF	45600	21000	Fluctuation_significant	microscopic	✓	USAF_dynamic_significant
11	Bottle	80000	27688	Constant	macroscopic		Bottle_static_constant
12	Bottle	80000	32258	Constant	macroscopic	✓	Bottle_dynamic_constant
13	Cardiac Muscle Section	30400	13025	Constant	microscopic		Cardiac_static_constant
14	Cardiac Muscle Section	30400	11100	Constant	microscopic	✓	Cardiac_dynamic_constant
15	Lily Anther Section	35600	14480	Constant	microscopic		Lily_static_constant
16	Lily Anther Section	35600	14000	Constant	microscopic	✓	Lily_dynamic_constant
17	Fan	20000	7949	Constant	macroscopic		Fan_static_constant
18	Fan	20000	12288	Constant	macroscopic	✓	Fan_dynamic_constant

Table 1. Details of the spike-based autofocus dataset

Scene Name	Mean	LPS [3]	Kurt [3]	HE [3]	NCES [3]	NCEB [4]	Tenegrad [5]	VF5 [6]	MF_DCT [1]	SD	SGFS
USAF_static_constant	-16640	24160	9880	580	-50	-50	23350	-140	160	-140	-29
USAF_static_decrease	19181	19481	18791	2981	-679	-679	-25819	-499	1211	161	-146
USAF_static_increase	-32680	12800	-13840	-5350	-70	-70	12800	-1270	-1450	-100	-62
USAF_static_minor	-28490	14770	14080	340	160	160	280	-110	-15290	280	-9
USAF_static_significant	-24360	20550	-22470	-21090	-2040	-2040	-2370	480	-21090	-1920	-149
USAF_dynamic_constant	-22600	13970	-14530	-5230	80	80	80	-490	-130	20	40
USAF_dynamic_decrease	22360	22480	13270	8170	-20	-20	-23090	-500	730	10	-19
USAF_dynamic_increase	-21010	15920	-11290	-20380	15770	15770	15830	-160	-19960	380	50
USAF_dynamic_minor	-17667	19803	-13437	-11337	273	273	333	243	-11517	243	54
USAF_dynamic_significant	-20520	20970	-21960	-20520	-1170	-1170	-1620	300	-20520	-1620	9251
Bottle_static_constant	25088	27488	-51312	26088	-512	88	88	-6512	-112	-512	-13
Bottle_dynamic_constant	31858	32058	-47342	31858	-742	-742	258	-5342	-742	-142	33
Cardiac_static_constant	-15165	13015	-15425	595	595	595	905	595	555	595	78
Cardiac_dynamic_constant	-19040	11090	-19210	-19040	-120	-120	-18340	-120	-18840	-180	-63
Lily_static_constant	-19680	14440	13680	-760	320	320	920	-80	-19680	40	-701
Lily_dynamic_constant	-14920	13960	-19440	-1120	640	640	1640	1760	-14920	-240	3774
Fan_static_constant	7749	7849	-11651	2649	-51	-51	-51	-51	7549	149	-31
Fan_dynamic_constant	8388	12188	-7612	1288	488	488	288	188	8388	488	-21

Table 2. Quantitative comparisons in the spike of each sequence with the relative error. A set of frame-based autofocus methods was applied to images reconstructed from spike using TFP [7] and the results in SGFS were directly derived from the extension of EGS to spike stream. The error is measured in frames, where, for instance, a value of 68 signifies a discrepancy between the manually pre-determined optimal focus position (focal frame as shown Tab. 1) and the frame predicted by the autofocus search. Due to the high temporal resolution of the spike frames, *e.g.* 50 s, the errors within 1000 are practically imperceptible to the human eye. In some scenarios, differences may not even be discernible within larger error margins, such as in the case of the Cardiac Muscle Section.

1. Appendix

In the appendix, we demonstrate why the power measure is effective in capturing high-frequency information of intensity, illustrating its capability to reflect the degree of defocus.

1.1. Analysis of Defocusing and Focusing

Lens systems are often simplified as thin lenses, and the imaging model adheres to the following equation:

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

where u is the distance between the lens plane to the object plane, v is the distance between the lens plane to the image plane, and f is the focal length. Typically, when an object point is not in the focal position, it results in a blurred image on the sensor. This blurred image shares a similar shape with the lens aperture but with a scaling factor. The aperture is commonly assumed to be circular with a diameter of D, and the circular blurred image formed on the imaging sensor is referred to as a blur circle. If we denote its diameter as R, the scaling factor is represented as K and can be defined as:

$$K=\frac{2R}{D}=\frac{s-v}{v}=s\left[\frac{1}{v}-\frac{1}{s}\right] \tag{2}$$

where s is the distance between the lens plane and the sensor plane, and it can be either greater than v or less than v. The former implies that the sensor is behind the focal plane, while the latter implies that the sensor is in front of the focal plane. Combining with Eq. (1), we can obtain:

$$R(u,s)=K\frac{D}{2}=s\frac{D}{2}\left[\frac{1}{f}-\frac{1}{u}-\frac{1}{s}\right] \tag{3}$$

Assuming the optical system is lossless, meaning there is no energy loss during the transmission of light, a blurred circle formed by a unit light source with a diameter of R is commonly referred to as the Point Spread Function (PSF). It can be represented as:

$$h_a(x,y)=\begin{cases}\frac{1}{\pi R^2} & \text{if } x^2+y^2\le R^2\\ 0 & \text{otherwise}\end{cases} \tag{4}$$

where $h_a(x,y)$ is the PSF, and the corresponding Optical Transfer Function (OTF) [8] can be expressed as:

$$H_a(\omega,\nu;u,s)=2\frac{J_1(R(u,s)\rho(\omega,\nu))}{R(u,s)\rho(\omega,\nu)} \tag{5}$$

where ω and ν are the spatial frequencies, J_1 is the first order Bessel function, and $\rho(\omega,\nu)=\sqrt{\omega^2+\nu^2}$ is the radial spatial frequency.During the process of imaging, it is akin to convolving the focused image $f(x,y)$ with the corresponding PSF. In the frequency domain, this can be expressed as:

$$G(\omega,\nu)=H_a(\omega,\nu;u,s)F(\omega,\nu) \tag{6}$$

From [8], we know that the OTF typically resembles a low-pass filter, influenced by the diameter R of the blur circle, where a larger R leads to a faster decrease in the main lobe area of the OTF, indicating more loss of high-frequency information. The increase in R essentially corresponds to an increase in defocus. Therefore, the area of the main lobe in the OTF can reflect the degree of defocus. The obvious approach is to directly calculate the area of this region as a focus measure, which is the classical power measure, defined as:

$$M_p=\iint|G(\omega,\nu)|^2\,d\omega d\nu \tag{7}$$

According to the Parseval's theorem, a more efficient computational approach can be defined as:

$$M_f=\iint|g(x,y)|^2\,dxdy \tag{8}$$

For practical numerical computations, ensuring numerical stability and avoiding issues such as overflow, an equivalent approach is::

$$M_1 = \frac{1}{A} \iint \left(g(x,y) - \mu_a\right)^2 dxdy \tag{9}$$

where μ_a is the mean of g, A is the area of g and the discrete form is:

$$M_d = \frac{1}{N} \sum_{\boldsymbol{x} \in \Omega} (G(\boldsymbol{x}) - \mu_d)^2 \tag{10}$$

where $G(\boldsymbol{x})$ is the pixel value a position $\boldsymbol{x}$ in the image, and N is the number of pixels in the image. Thus far, the degree of defocus can be reflected through M_d. As indicated in Eq. (3), it is evident that the variable R is predominantly influenced by u and s. In macroscopic imaging, u is typically held constant, and achieving focus involves adjusting s. Conversely, in microscopic imaging, particularly within a 4f system, s is often fixed, and achieving focus is accomplished by manipulating u. Despite this distinction, the fundamental adjustment of these parameters boils down to modifying the size of R which is tantamount to altering the PSF of the system.

1.2. Complexity Analys

The complexity of focus measure SD is $\mathcal{O}(n * m * w)$, where n is the width of the sensor's pixels, m is the height of the sensor's pixels and w is the length of the integration window. Leveraging parallel optimisation can further reduce this complexity to $\mathcal{O}(w)$. Generally, the autofocus traversal process can be extremely rapid, exemplified by the spike camera mentioned in [9], which captures 40,000 spike frames in 1s. As indicated in [10], computations under complexity $\mathcal{O}(w)$ are very efficient, reaching a microsecond level when w is less than 1×10^6. This can meet the real-time feedback calculation requirements in SGFS. Moreover, due to the shortened traversal range, the reduction in the number of iterations in SGFS can further reduce the time consumption, compared to EGS [10].

1.3. Cases Analysis

In the main text, we presented the results of simulated and real-world data under varying light-intensity conditions. In the supplementary materials, we further augmented the results with additional scenarios of light intensity variations Figs. 1 and 2, and additional real-world data Figs. 3 and 4. The quantitative comparisons in spike with the relative error of each sequence are shown in Tab. 2. All these results consistently demonstrate that SD exhibits stronger robustness to changes in light intensity, whereas other methods are susceptible to varying degrees of impact. Additionally, We have also found that, due to spike noise, previous frame-based methods struggle to achieve rapid autofocus in high-speed conditions, resulting in significant errors.

Additionally, we have observed that in scenarios with lower light intensity, the presence of heightened noise significantly interferes with other methods. In contrast, SD exhibits a superior ability to withstand the influence of noise, as depicted in Fig. 3(a).In real-world scenarios, the presence of noise can affect the precision of autofocus.

From Tab. 2, we observe that the SGFS method exhibits a relatively larger error in $Lily\_dynamic\_constant$(3774), $Lily\_static\_constant$(701), $USAF\_dynamic\_significant$(9251). For $USAF\_dynamic\ significant$, it is primarily because the SGFS method employs a large accumulation window, which violates the assumption of constant light intensity when encountering significant fluctuations in light intensity. However, given the inherently robust performance of SD against light intensity disturbances, it is viable to directly utilize the results from SD for peak searching in this scenario. For $Lily\_dynamic\_constant$ and $Lily\_static\_constant$, the primary reason is the limited structural information available in the samples, coupled with the slow temporal changes in the focusing process. As a result, even with significant temporal errors, the actual displacement distance is relatively short, and the results still fall within the depth of field range.

References

[1] 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. 1, 2, 3, 6

[2] 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. 1, 2, 3

[3] 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. 6

[4] 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. 6

[5] TTE Yeo, SH Ong, Jayasooriah, and R Sinniah. Autofocusing for tissue microscopy. *Image and Vision Computing*, 11(10):629–639, 1993. 6

[6] 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. 6

[7] 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. 6

[8] Murali Subbarao, Tae-Sun Choi, and Arman Nikzad. Focusing techniques. In *Machine Vision Applications, Architectures, and Systems Integration*, volume 1823, pages 163 – 174, 1992. 7

[9] 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. 8

[10] 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. 8