Optics and Lasers in Engineering 160 (2023) 107247

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Millisecond autofocusing microscopy using neuromorphic event sensing

Zhou Ge[a,*], Haoyu Wei[a], Feng Xu[a], Yizhao Gao[a], Zhiqin Chu[a,b], Hayden K.-H. So[a,c], Edmund Y. Lam[a,c]

[a] Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[b] School of Biomedical Sciences, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[c] AI Chip Center for Emerging Smart Systems (ACCESS), Hong Kong Science Park, Hong Kong SAR, China

ARTICLE INFO

Keywords:
Autofocusing
Microscopy
Event sensing

ABSTRACT

Rapid autofocusing is essential for many microscopic imaging applications. Existing methods either require complicated hardware implementations or slow z-stack image acquisition. In this work, we develop a new approach to achieve fast autofocusing using the neuromorphic event sensing technology, which detects the sparse brightness changes asynchronously and responds quickly to the specimen movement. A simple yet efficient autofocusing system is tailored for fast acquisition and processing of the non-redundant event data, allowing for autofocusing in only tens of milliseconds, which is thousands of times faster than current technologies. Experimental results show a substantial performance improvement and capability for biopsy specimen inspections. To the best of our knowledge, this is the first reported work on a neuromorphic methodology for autofocusing microscopy.

1. Introduction

Optical microscopy is foundational in many areas of clinical diagnosis and biomedical research. Due to the shallow depth of field (DOF) in microscopy systems, rapid and accurate autofocusing is essential in order to capture sharp images for the subsequent analysis. Nevertheless, it is a challenging process for many microscopy applications, such as whole slide imaging [1,2], light sheet microscopy [3] and high content screening [4], due to the stringent requirements on speed and precision.

Existing autofocusing methods can be broadly categorized as hardware methods and software methods. Hardware methods directly correct the changes in the focal plane by measuring the distance between the objective lens and the specimen. They tend to have a high focusing accuracy and rapid response, but generally require expensive hardware modifications, such as infrared laser interferometry or confocal pinhole detection [5,6]. Consequently, software methods are more preferred for practical applications. They determine the optimal focus position by calculating the sharpness of a set of differently focused images [7–9]. While these methods are rather flexible, they are more time-consuming due to the heavy computation and the speed is fundamentally limited by the frame rate of the imaging sensors.

Inspired by biological visual systems, neuromorphic event sensors are a new type of electronic sensors where each photodetector works asynchronously and responds only to brightness changes as ON and OFF events. This brings many advantages, including high temporal precision, low power consumption, wide dynamic range, and redundancy reduction [10,11]. The neuromorphic event sensing has facilitated many innovative applications recently, such as depth estimation [12,13], motion sensing [14–16], and gesture recognition [17]. This technology is also attractive for microscopy applications such as neural imaging [18] and microparticle tracking [19,20].

In this work, we demonstrate the usage of the neuromorphic event sensing to achieve rapid autofocusing. To the best of our knowledge, this is the first report showcasing the application of event sensors in autofocusing microscopy. The rest of the paper is organized as follows: in Section 2, we review the existing autofocusing methods in microscopy and commonly used focus metrics. Then, in Section 3, we explain the details of our approach, including the mathematical derivation of the wavefront distribution and the event generation model. In Section 4, we present the experimental setup and analyze the performance of our proposed method in rapid scanning and real biopsy specimens. Based on the above, we draw the conclusions in Section 6.

2. Existing methods

2.1. Rapid autofocusing microscopy

Traditionally, autofocusing is achieved using what can be known as focus map surveying [21]. By capturing a stack of images from an axial scanning and assessing the degree of focus of each image, the focus position can be determined by maximizing a certain focus metric. (Common choices of the metric will be discussed in the next section.) This method is widely applied in many commercial automated microscopy systems

* Corresponding author.
E-mail addresses: gezhou@eee.hku.hk (Z. Ge), elam@eee.hku.hk (E.Y. Lam).

https://doi.org/10.1016/j.optlaseng.2022.107247
Received 31 May 2022; Received in revised form 27 July 2022; Accepted 29 August 2022

Table 1
Commercial autofocusing microscopy systems [21].

Vendor	Sensor	Method
Leica	TDI	Focus map/Dual sensor [32]
Zeiss	CCD	Focus map
Nikon	CCD	Triangulation [33]
Olympus	CMOS	Focus map/Triangulation

due to the simplicity in hardware and robustness against various conditions. However, a major hurdle is the trade-off between speed and accuracy. To cover the correct focus position in the z-stack, we need to have a large sample size at different depths. Consequently, it can take many seconds or even up to several minutes for a conventional video-rate camera to complete the procedure.

Alternatively, reflective-based methods have been proposed to improve the autofocusing speed. They typically do not require a complete z-stack, thus significantly reducing the time for data acquisition. Instead, they detect a reference plane to locate the sample plane rapidly. One way of localizing the reference plane is through a confocal pinhole detection, where the refractive index differences in the sample space is used to identify a strong peak during an axial search. Then, a refinement search for the precise focus position can be performed within a small neighborhood. There are more efficient methods, such as triangulation and low-coherence interferometry [22–24]. However, they often require a complicated hardware system or an active illumination, which can be a burden for many applications.

Another line of work accelerates image-based methods by using a separate focusing sensor. Traditionally, the same imaging sensor is used for both focusing and image acquisition, which potentially limits the performance. A dual-sensor system uses a separate sensor for the focus survey where high-resolution images are captured by the main imaging sensor once the focus position is determined by the focusing sensor. A beam splitter array or phase modulation induced by some coded masks can further integrate the different focus information through a single-shot acquisition [25–27]. However, a higher level of multiplexing inevitably reduces the autofocusing accuracy, thus this approach is applicable only to a short focal range. Furthermore, these methods require careful calibrations and rely heavily on the quality of captured images.

In recent years, learning-based methods have also been explored for autofocusing [28–31]. While they are shown to be rather effective to either predict the defocus distance or refocus the blurry image, a considerable amount of data is generally required for training the neural networks. The algorithms may also fail for new features that are out of the training distribution.

The sensors and methods used in commercial autofocusing microscopy systems are listed in Table 1. Focus map surveying is the most widely used method due to its robustness and universality. Besides the commonly used charge-coupled device (CCD) and complementary metal oxide semiconductor (CMOS) image sensors, time delay integration (TDI) sensor is also adopted with its better performance in photon-limited applications.

2.2. Focus metrics

The principle of z-stack surveying is shown in Fig. 1. The sample is axially scanned at different depths, and a camera is used to acquire the z-stack images. When the sample is in focus, the image would give sharper edges, larger contrasts, and a larger intensity variance. When the sample departs from the focus position, convolution with the defocus point spread function (PSF) blurs the image. This attenuates the high frequency components of the images and critical details are inevitably lost. We can therefore design various measures, so that the best score is attained for an image that is sharply in focus.

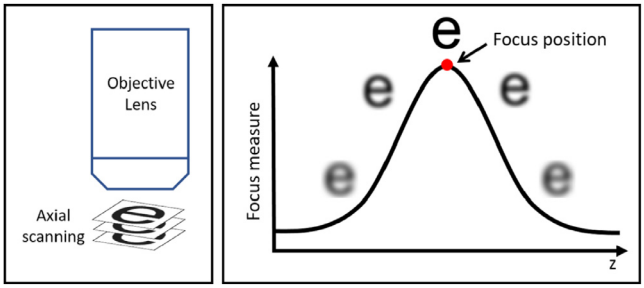


Fig. 1. (a) The sample is axially scanned at different focal depths, and a z-stack of images is obtained. (b) The focal plane can be determined by maximizing a certain focus metric.

Over the years, many focus metrics have been proposed. They can generally be divided into four groups: gradient methods, statistical methods, histogram methods, and intuitive methods [34]. In what follows, we select one from each group [35–38]:

$$(\ell_1\text{-norm of the gradient}) \quad M_1 = \sum_{x,y} \Big(|u_x| + |u_y| \Big), \tag{1}$$

$$(\text{image variance}) \quad M_2 = \sum_{x,y} \Big(I(x,y) - \mu \Big)^2, \tag{2}$$

$$(\text{histogram entropy}) \quad M_3 = -\sum_{i=1}^{l} p_i \log_2 \left(p_i \right), \tag{3}$$

$$(\text{image power}) \quad M_4 = \sum_{x,y} \Big(I(x,y) \Big)^2. \tag{4}$$

In these equations, u_x and u_y stand for the image gradient at pixel (x, y), μ denotes the mean intensity of $I(x, y)$, p_i is the probability of a pixel with intensity value i, and l is the number of gray levels.

Besides these focus measures, we also quantitatively evaluate the performance from two aspects: range and reproducibility [9]. Range represents the length of the extreme value. It is given by the full width at half maximum (FWHM). Reproducibility is computed as the width at 80% of the extreme value. A smaller reproducibility corresponds to a sharper peak in the statistics.

3. Neuromorphic event-based autofocusing

3.1. Principle of neuromorphic event sensors

The working principle of a neuromorphic event sensor is presented in Fig. 2. Information of the scene is transmitted as ON and OFF events when the brightness changes exceed the predefined threshold. The mathematical representation of the event generation process is given as

$$|L(x,y,t) - L(x,y,t-\Delta t)| \geq C, \tag{5}$$

where L represents the brightness signal which can be interpreted as logarithmic intensity in uniformly-illuminated scenes, t denotes the timestamp, Δt is the time elapsed since the previous event triggered at the same pixel, and C is the predefined threshold. The sign of the brightness change is recorded using a binary polarity p.

Due to the unique design of signal processing circuitry, the event sensors can monitor fast changes in the signals with microsecond resolution, i.e., Δt is in the order of microseconds. Moreover, each sensor pixel works independently and asynchronously. An output is instantly triggered and transmitted once a pixel detects a sufficient brightness change, so there is no need to use a shutter or set an exposure time. The

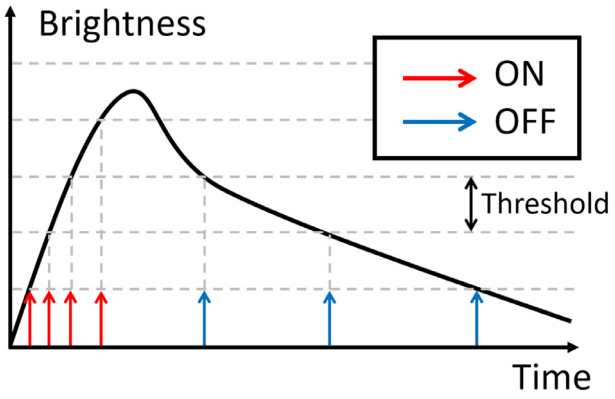


Fig. 2. The working principle of the neuromorphic event sensor. A continuous brightness (logarithmic intensity) signal is converted as a sequence of ON and OFF events.

event sensor sends a stream of continuous output and each of them encodes the pixel location, timestamp, and polarity into a four-dimensional vector, which can be mathematically represented as

$$e_s = (x_s, y_s, p_s, t_s), \tag{6}$$

where s is the index of the event stream.

New methods should be developed to extract information from the event stream due to its unique data structure. Existing event data processing algorithms can be broadly categorized into two classes. The first group of algorithms operate on an event-by-event basis, which updates the current state according to the mismatch information from each incoming event [39,40]. These methods exploit the asynchronous attribute of the event stream and can process real-time tasks with minimal latency. However, to do so, they require some additional information, such as grayscale images or a global map. The second class of algorithms aggregate a group of events into an event frame. This representation is then compatible with existing image-based methods, such as machine learning with a support vector machine (SVM) and deep learning with a convolutional neural network (CNN) [41,42]. These methods generally have a high signal-to noise ratio (SNR), at the expense of losing temporal correlation.

3.2. Wave distribution in the axial direction

In order to harness the neuromorphic event sensing for autofocusing, we first determine how the detected brightness at the sensor plane changes as a result of the sample's movement in the axial shifting. We consider an imaging system that consists of a single convex lens with a focal length f. An on-axis unit impulse of wavelength λ is positioned at a distance z_1 from the center of the lens on one side, and the sensor is positioned at a distance z_2 on the other side. The unit impulse produces a spherical wave before the lens, with a paraxial approximation written as [43]

$$U\left(x_a, y_a; z_1, 0^-\right) = \frac{1}{j\lambda z_1} \exp\left[\frac{jk}{2z_1}\left(x_a^2 + y_a^2\right)\right], \tag{7}$$

where $k = 2\pi/\lambda$ is the wave number and (x_a, y_a) denotes the aperture plane coordinates. The convex lens performs a phase transformation to that wave and induces a modulated converging spherical wave after the lens, which is given by

$$U(x_a, y_a; z_1, 0^+) = \frac{1}{j\lambda z_1} A(x_a, y_a) \exp\left[\frac{jk}{2}\left(\frac{1}{z_1} - \frac{1}{f}\right)\left(x_a^2 + y_a^2\right)\right], \tag{8}$$

where $A(x_a, y_a) = 1$ for a point (x_a, y_a) inside the aperture, and 0 otherwise. Finally, the field propagates a distance z_2 to the sensor. By Fresnel diffraction approximation, the amplitude complex field arriving at the sensor plane (x, y) can be described as

$$\begin{aligned} U(x, y; z_1, z_2) &= \iint_{-\infty}^{+\infty} U(x_a, y_a; z_1, 0^+) \exp\left[\tfrac{jk}{2z_2}\left(x_a^2 + y_a^2\right)\right] \\ &\quad \times \exp\left[\tfrac{jk}{z_2}\left(x x_a + y y_a\right)\right] \mathrm{d}x_a \mathrm{d}y_a \\ &= \iint_{-\infty}^{+\infty} A(x_a, y_a) \exp\left[\tfrac{jk\beta}{2}\left(x_a^2 + y_a^2\right)\right] \\ &\quad \times \exp\left[\tfrac{jk}{z_2}\left(x x_a + y y_a\right)\right] \mathrm{d}x_a \mathrm{d}y_a, \end{aligned} \tag{9}$$

where $\beta = \left(\frac{1}{z_1} + \frac{1}{z_2} - \frac{1}{f}\right)$. It can be seen that we have ignored the phase variation that precedes the integral, since we only consider its intensity distribution. In order to measure the brightness changes when the object is axially shifted, we further simplify the equation by setting $x = y = 0$, since most of its energy is concentrated at the center position, and the central magnitude effectively models the major source of the brightness changes that can be detected by an event sensor. At $x = y = 0$, we obtain

$$U(0, 0; z_1, z_2) = \iint_{-\infty}^{+\infty} A(x_a, y_a) \exp\left[\frac{jk\beta}{2}\left(x_a^2 + y_a^2\right)\right] \mathrm{d}x_a \mathrm{d}y_a. \tag{10}$$

By converting into polar coordinates, namely $x_a = \rho\cos(\phi)$ and $y_a = \rho\sin(\phi)$, we rewrite the relationship between the source and destination amplitude fields with respect to the depth z_1 and z_2 as

$$\begin{aligned} U(0, 0; z_1, z_2) &= \int_0^{2\pi} \int_0^a \exp\left[\tfrac{jk\beta\rho^2}{2}\right] \rho \mathrm{d}\rho \mathrm{d}\phi \\ &\propto \frac{\sin\left(\frac{\pi\beta a^2}{2\lambda}\right)}{\frac{\pi\beta a^2}{2\lambda}} = sinc\left(\tfrac{\beta a^2}{2\lambda}\right), \end{aligned} \tag{11}$$

where a is the radius of the aperture. Hence, for a microscopy system, the axial wave field follows a sinc distribution with a different magnification factor. Assuming that the PSF is shift-invariant, an off-axis point source $(u, v \neq 0)$ results in the same pattern as the on-axis one of the same depth, but centers at its corresponding image point.

3.3. Autofocusing from an asynchronous event stream

The next step is to determine the focus position from the event data. During the axial scanning procedure, the sample is axially shifted, and the events are monitored at a fixed image plane. Hence, z_2 is a fixed parameter in Eq. (11). For simplicity, we use z to refer to the variable z_1 in the following equations to represent the depth of the sample. The detected brightness signal can be mathematically modeled by

$$L(x, y; z) = \log\left(f(x, y) \otimes h(x, y; z)\right), \tag{12}$$

where $L(x, y; z)$ denotes the detected brightness at location (x, y) of the sensor with respect to the sample at depth z, while $f(x, y)$ denotes the 2D intensity value of the sample, $h(x, y; z)$ represents the PSF and $\otimes$ stands for the convolution operator. The PSF can further be derived from the wave distribution as

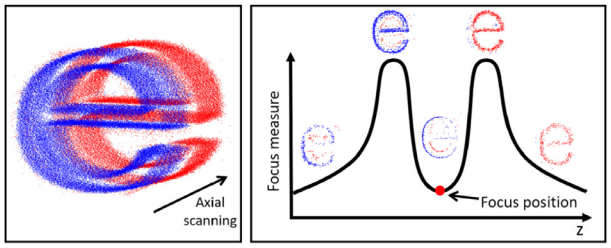
$$h(x, y; z) = \left|U(x, y; z)\right|^2. \tag{13}$$

Assuming that the axial movement of the sample is linearly related to time by a velocity v (i.e., $z = v \cdot t$), from Eq. (5), we can obtain the relationship between the event generation and the sample's position as

$$|L(x, y; z) - L(x, y; z - \Delta z)| \geq C, \tag{14}$$

where Δz represents a small axial displacement of the sample. An event is triggered if the above inequality is satisfied. By taking a sufficiently small displacement Δz, the number of events at each axial position z is linearly related to the absolute value of the brightness gradient and is given by

$$\left|\frac{L(x, y; z) - L(x, y; z - \Delta z)}{\Delta z}\right| \approx \left|\frac{\partial L(x, y; z)}{\partial z}\right|. \tag{15}$$



(a) (b)

Fig. 3. (a) During the axial scanning, events are mainly triggered symmetrically at both sides near the focal plane and almost undetectable at the in-focus position. (b) The proposed method gives a focus measure that follows an "M" shape, and the focus position is located at the valley between the peaks.

In order to demonstrate the relationship between the depth and the generated events, a scatter plot of the events is shown in Fig. 3 a where a sample with an opaque letter "e" on a transparent background is axially scanned using bright-field microscopy. The red pixels represent ON events, and the blue pixels represent OFF events. In bright-field microscopy, the background has a higher brightness level than the illuminated sample. Therefore, when the sample is shifted from an out-of-focus position to an in-focus position, the brightness decreases at the edges of the pattern and OFF events are generated in this process. Similarly, ON events are generated when the sample is moving away from its in-focus position. Since the wave distribution can be modeled as a sinc function as given in Eq. (11), the brightness maintains its highest value at the central peak. Consequently, events are almost undetectable when the sample moves to the in-focus position, i.e., the interval between the bunches of ON and OFF events in the scatter plot. Such variations in the event generation can be utilized to indicate the focal plane effectively.

In our method, the recorded data stream is processed by accumulating the events generated between z and $z+\Delta z$ irrespective of their polarities. The depth is transformed to the time domain according to the known velocity, giving rise to the metric

$$M_5 = \sum_{x_s,y_s} \sum_{z_s=z}^{t+\Delta t} [e_s], \tag{16}$$

where $[e_s]$ denotes the number of triggered events expressed in Eq. (6). As shown in Fig. 3 b, the curve describes the theoretical autofocusing measurement with the proposed metric. The proposed method gives a focus measure that follows an "M" shape, and the focus position is located at the valley between the peaks. The distance between the two peaks is about a few dozens of micrometers depending on the numerical aperture (NA) of the objective lens, tube length and scene geometry. In practical experiments, the scanning distance may be contained in the interval between the two peaks and results in a "V" curve in the numerical plot. Under such a circumstance, the proposed method can be regarded as unimodal, with the minimum of the focus measure indicating the in-focus position.

We also plot several event frames for a better visual inspection of the event generation at different depths. When the sample is far away from the focus position, the sample is completely out-of-focus and axial movements cause very few brightness changes at each pixel, hence very few events are monitored. The detected brightness changes significantly when the sample moves near the focus position and reaches to its maximum at position z^* that can be estimated by

$$\begin{aligned} z^* &= \arg\max_z \left| \frac{\partial L(x,y;z)}{\partial z} \right| \\ &= \arg\max_z \left| \frac{\partial \log\left(f(x,y) \otimes h(x,y;z)\right)}{\partial z} \right| \\ &\approx \arg\max_z \left| \frac{\partial \log |U(0,0;z)|^2}{\partial z} \right|, \end{aligned} \tag{17}$$

where $U(0,0;z)$ can be modelled as $sinc(\frac{\beta a^2}{2\lambda})$ according to Eq. (11).

Table 2
Technical parameters of the APS and the DVS in the DAVIS346 neuromorphic event sensor.

Parameter	APS	DVS
Pixel size	18.5 μm × 18.5 μm	18.5 μm × 18.5 μm
Pixel number	346 × 260	346 × 260
Temporal resolution	25 ms	1 μs
Dynamic range	56.7 dB	120 dB

4. Experiments and results

4.1. Experimental setup

The experimental setup is shown in Fig. 4 a. The microscope (Eclipse Ni-U, Nikon) uses a 100W Halogen lamp as the light source. The sample is axially shifted using a motorized linear translation stage (WN262TA20, Winner Optics) and the data is captured using a neuromorphic event sensor (DAVIS346, iniVation). The DAVIS event sensor integrates a conventional image-based active pixel sensor (APS) and an event-based dynamic vision sensor (DVS) [44], so that in the meantime of recording event data, image frames can be captured as well. The technical parameters of the APS and the DVS are summarized in Table 2. In the experiments, we use the DVS event sensor to test our proposed method and use the APS sensor to capture image frames as comparison.

To show a real example of the APS output and DVS output of the event sensor, we capture a moving pollen specimen with a 10×/0.25 objective lens. The APS captures the absolute intensity value in the whole field of view (FoV) as an 8-bit RGB image frame, which is shown at the top of Fig. 4 b. Both the pollen and the background are clearly captured in the scene. The asynchronous event stream is reconstructed into a 2D event frame by accumulating the events within a 10 ms time window, shown at the bottom of Fig. 4 b. The reconstructed event frame clearly shows the outline of the pollen as ON and OFF events and filters out the background. For a rough comparison of their data efficiency, the image frame encodes the scene information using the full pixels, resulting in $346 \times 260 = 89,960$ values being recorded. On the other hand, the event frame only consists of 1435 events in this case since only the pixels that detect brightness changes transmit information.

4.2. Image-based autofocusing

First, we demonstrate the experimental results using the conventional image-based autofocusing methods. The test sample is a resolution target (HIGHRES-2, Newport), which has opaque lines on a clear background. We use a 40×/0.65 objective lens with a 2 μm DOF in the experiment. The sample is axially shifted for 50 μm with a step size of 1 μm. Twelve focal images taken at a step size of 4 μm are presented in Fig. 5 a. As the position of the sample shifts along the depth axis, the focal images turn from blurry to sharp, and blurry again. Since the conventional imaging sensors could only capture clear images of static scenes, the moving stage spends a significant amount of time pausing, accelerating and decelerating for each image frame. As a result, it takes over one minute to acquire the z-stack of 50 images using the APS with the exposure time of 1 ms. Besides, data redundancy could be a problem. A sequence of 50 image frames of 346×260 resolution takes 1.24 MB to store in 8-bit RGB format, most of which are repetitive information.

Using our naked eyes, it's difficult to determine which image is best focused in Fig. 5 a since they have very slight differences. Hence we numerically analyze the whole 50 z-stack images using the focus measures in Eqs. (1)–(4). The results are normalized within the range of $[0, 1]$ and are shown in Fig. 5 b. The histogram entropy method fails in predicting the correct focus position due to the lack of variety in sample intensities. All the other three focus measures indicate that the focus position is at about 19 μm. The mean reproducibility is 0.58 and the range is 0.83. We also record the computation time for the

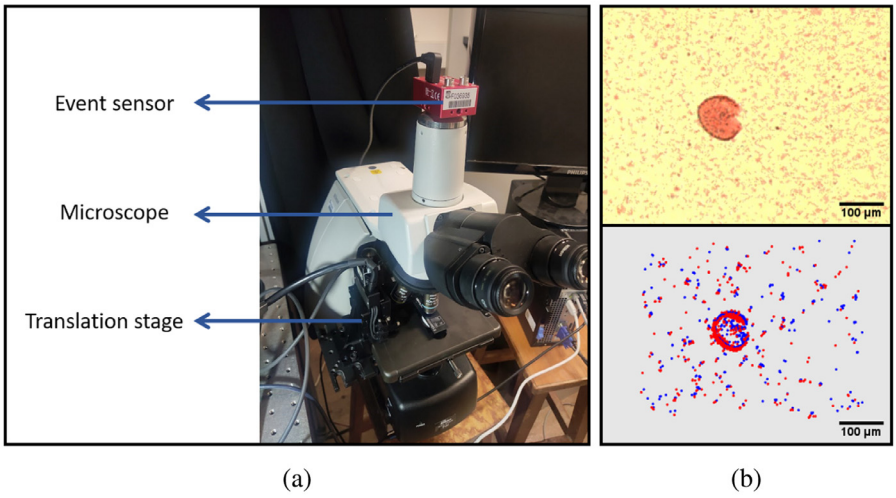


Fig. 4. (a) Experimental setup of the autofocusing bright-field microscope. (b) APS (top) captures the absolute intensity values of the whole FoV while DVS (bottom) only records the brightness changes on the edges.

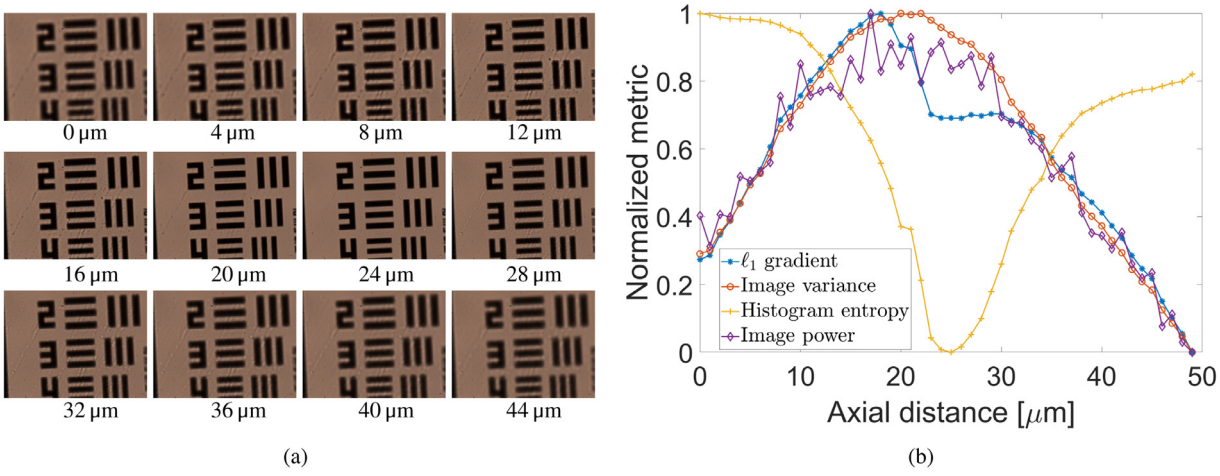


Fig. 5. (a) A z-stack of images is obtained when the sample is shifted at different focal depths. (b) Numerical results using different focus metrics.

four autofocusing methods. The average computation time is 342.3 ms on a computer with an Intel i5-6500@3.20GHz CPU and MATLAB R2020a.

4.3. Event-based autofocusing

Then, we perform the same set of experiments on our proposed event-based autofocusing method. In this experiment, we continuously shift the sample at a uniform speed of 5 mm/s and record the data using the DVS. Our method allows for the continuous shifting of the sample, which saves a major amount of time in data acquisition. For the same moving distance of 50 μm, it takes only 10 ms to acquire all of the event data. Unless otherwise specified, we use the same scanning speed and distance in the following experiments. A total of 53,274 events are recorded during the whole process, and they consume only 130 KB for storage. It is a substantial improvement compared with the one-minute data acquisition time and 1.24 MB storage consumption using the image-based autofocusing methods.

Although our method works directly on the asynchronous event data and does not depend on the frame representation, we transform them into event frames for better visualization and comparison with the conventional methods. Unlike ordinary imaging sensors, where the frame rate cannot be adjusted after the data acquisition, the number of event frames can be user-defined after the data acquisition by setting different integration window time. As such, we reconstruct the event frames into the same step size of 4 μm as the image-based autofocusing experiment by integrating the events in every 0.8 ms.

The z-stack of 12 event frames is presented in Fig. 6 a. Compared with the image frames in Fig. 5 a, each event frame has remarkable differences. As the position of the sample shifts along the depth axis, the recorded events gradually become sparse, and become dense again. From 0 μm to 20 μm, many negative events are detected since the opaque patterns become darker when the sample gets focused. From 20 μm to 44 μm, the patterns become brighter as the background light disperses to those regions, so that many positive events are detected. As previously discussed, the axial movement of the sample near the in-focus position will not cause significant brightness changes, hence very few events occur in this region. Using our naked eyes, we can easily assert the focus position is at the depth of 20 μm according to Fig. 6 a.

In order to achieve higher precision, we analyze the event data using the proposed algorithm in Eq. (16) and the numerical results are presented in Fig. 6 b after smoothing. The y-axis presents the number of events within every 50 μs integration time. As the moving speed is 5 mm/s, the step size in the x-axis is 0.25 μm. The experimental data is shown to be "V"-shaped since the scanning distance is within the range between the two peaks in Fig. 3 b. The minimizer is at 18.75 μm and

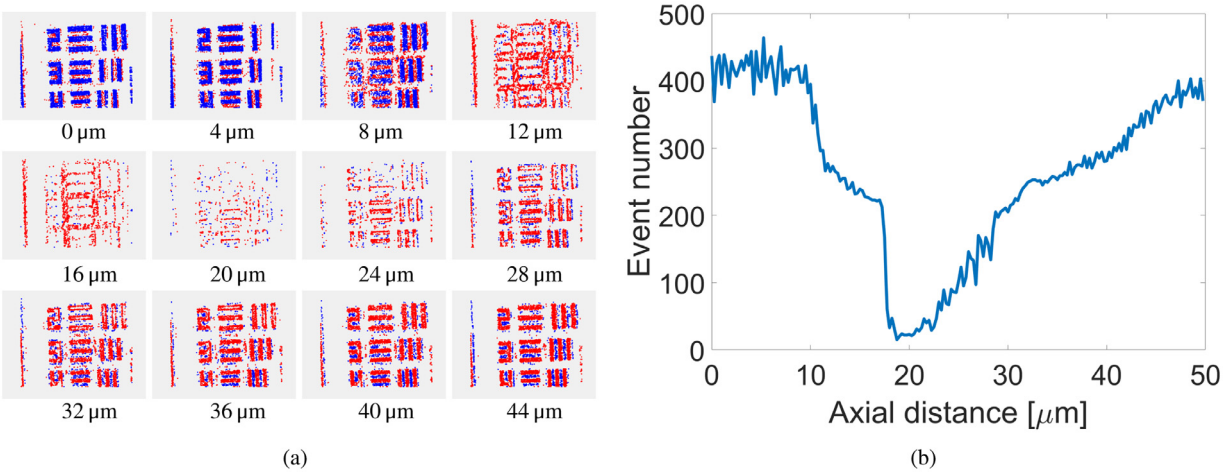


Fig. 6. (a) The z-stack event frames are reconstructed at different depths of the sample. (b) Numerical results using the proposed event-based autofocusing method.

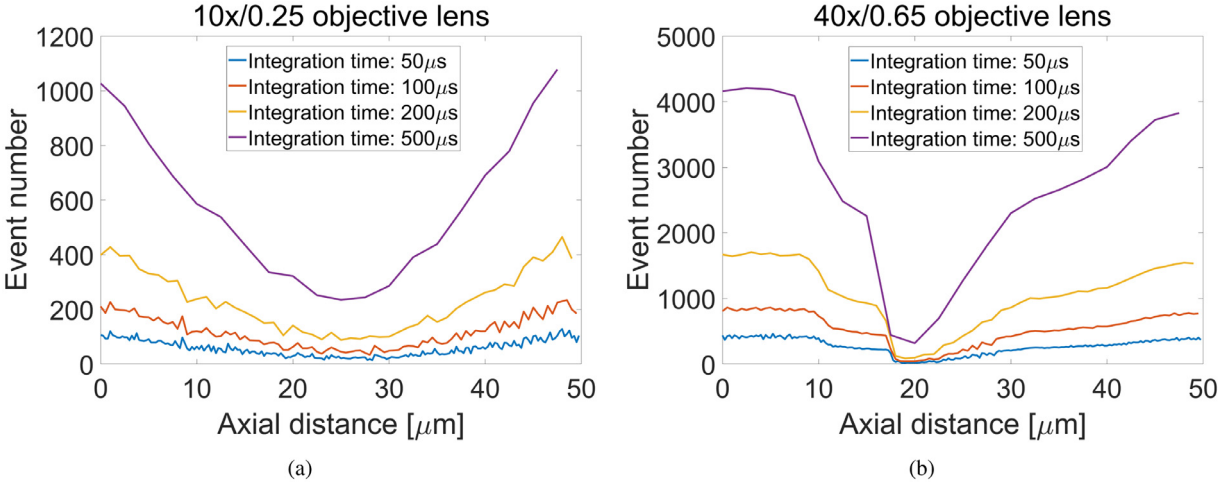


Fig. 7. Autofocusing results using a 10×/0.25 objective lens and a 40×/0.65 objective lens with different integration time settings.

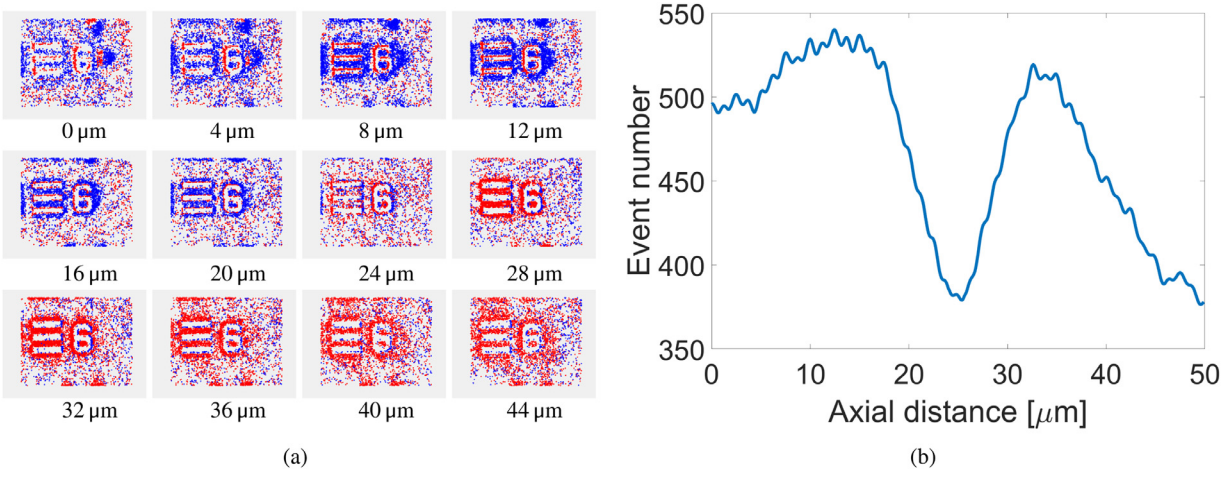


Fig. 8. (a) The z-stack event frames of the negative sample with a step size of 4 μm. (b) Numerical results using the proposed event-based autofocusing method.

this result accords with the visual inspection of the event frames and the image-based autofocusing methods. The reproducibility of the proposed method is 0.15 and the range is 0.295. The computation time is 4.2 ms on the same computer with an Intel i5-6500@3.20GHz CPU and MATLAB R2020a.

4.4. Performance comparison

The comparison between the event-based autofocusing and image-based autofocusing is summarized in Table 3. Image-based autofocusing methods rely on the intensity images, which cost more time to acquire

Fig. 9. The frame image of the negative sample at corresponding estimated depth of 25.25 μm.

Table 3
Comparison between the event-based autofocusing and image-based autofocusing.

	Image-based autofocusing	Event-based autofocusing
Data	50 images	53,274 events
Acquisition time	1 min	10 ms
Data storage	1.24 MB	130 KB
Processing time	226 ms–425 ms	4.2 ms
Precision	1 μm	0.25 μm
Reproducibility	0.58	0.15
Range	0.83	0.295

and process. In contrast, the proposed event-based autofocusing method significantly reduces both autofocusing time and storage consumption. Moreover, high precision can be achieved using the event sensing technology with microsecond temporal resolution, while image-based methods are physically limited to lower frame rates. The proposed method also outperforms the image-based methods in reproducibility and range, which demonstrates its superiority.

5. Discussion

5.1. NA and integration time

In the above experiments, we have shown the substantial performance improvement using the proposed event-based autofocusing. In applications such as whole slide imaging, a set of micrographs have to be captured using different objective lenses. Accordingly, the settings of the focusing camera, such as the frame rate and exposure, need to be changed due to the different NA and DOF. If the camera is not set properly, biased predictions will be made using the image-based autofocusing algorithms. However, the settings of the event sensors can be consistent when using different objective lenses, since the scene information is automatically monitored at a high temporal resolution. Moreover, the integration time is a user-defined parameter and can be adjusted after the data acquisition. Such characteristics provide convenience for practical applications.

We conduct comparative experiments using a 10×/0.25 objective lens and a 40×/0.65 objective lens. The DOF of these objective lenses are 16 μm and 2 μm, respectively. Theoretically, when using an objective lens with a large NA or using a long integration time setting, a large amount of events will be counted in each time window, and this results in the distinct numerical differences in the y-axis with higher SNR. However, it will also cause a compromise in the temporal resolution and reduce the precision. Hence, it is necessary to adjust to a proper integration time setting for different objective lenses to achieve the balance between SNR and precision.

We then test the proposed algorithm with different integration time ranging from 50 μs to 500 μs. The corresponding precision can be calculated as ranging from 2.5 μm to 0.25 μm. The results for the two tested objective lenses are shown in Fig. 7a and b, respectively. Comparing the two figures, the objective lens with a smaller NA has a flatter curve in the statistics and the focus position can only be distinguished using a 500 μs integration time with 2.5 μm precision. In comparison, the objective lens with a higher NA has a much sharper peak in the statistics and the focus position can be distinguished using a 50 μs integration time (see Fig. 6 b) with 0.25 μm precision. Hence, when using small NA objective lenses, a longer integration time can help to increase the

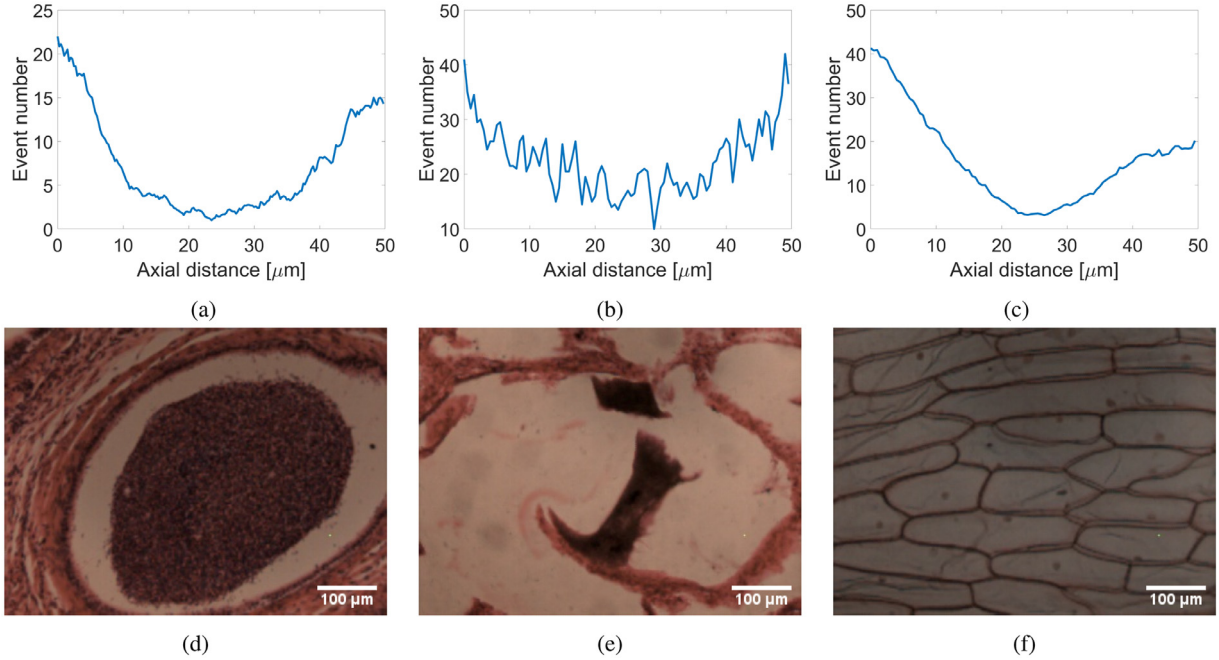


Fig. 10. (a)–(c): Autofocusing results of testis, pulmonary vessels, and onion cells. (d) - (f) Focal images taken at corresponding estimated depths.

SNR and a short integration time can be adopted to indicate the focus position with a higher accuracy when using high NA objective lenses.

5.2. Negative sample

In applications such as fluorescence microscopy, the sample has a higher brightness level than the background. Here, we simulate such a scenario using a negative sample (R3L3S1N, Thorlabs) with a 40×/0.65 objective lens. The z-stack of 12 event frames is presented in Fig. 8 a with a step size of 4 μm. The event frame with the sparsest events is around the depth of 24 μm from visual observation. Compared with the experimental results of a positive sample in Fig. 4 a, the background has a higher level of noise in the negative sample because the event sensor responds to logarithmic brightness changes and is more sensitive in the dark. The numerical results are presented in Fig. 8 b. The integration time is set to be 50 μs and the precision can be reached up to 0.25 μm. According to the metric in Eq. (16), the minimizer is at 25.25 μm, which is close to the visual inspection of the event frames. We also capture the frame image at the estimated axial distance in Fig. 9. The sharpness of the frame image confirms the accuracy of our method.

5.3. Biopsy specimen

Last but not least, we examine the validity of our method on real biopsy specimens including testis, pulmonary vessels, and onion cells using a 10×/0.25 objective lens with 16 μm DOF. The autofocusing results using the proposed method are shown in the upper row of Fig. 10, where the integration time is set to be 100 μs and the precision is 0.5 μm. The curve represents the changes in the detected event number with their minimum value indicating the focus position. The corresponding images captured at the indicated depths are shown in the bottom row of Fig. 10. They are all visually clear and sharp, which show the validity of our proposed method in biopsy specimen inspection. The processing time is less than 3 ms in all three cases, demonstrating the efficiency of our method.

6. Conclusion

In this work, we propose a new method to use the neuromorphic event sensing for rapid autofocusing in microscopy. The wave distribution in the axial direction and the event generation model during the autofocusing process are mathematically analyzed. Our method takes advantage of the high temporal resolution of the event sensor to achieve millisecond speed and sub-micrometer accuracy at the same time without using complicated hardware. Experimental results show the validity and efficiency of our method on biopsy specimens. Our work will be beneficial for fast-moving object observation and high throughput microscopy.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Zhou Ge: Methodology, Writing – original draft. **Haoyu Wei:** Methodology, Writing – review & editing. **Feng Xu:** Writing – review & editing. **Yizhao Gao:** Writing – review & editing. **Zhiqin Chu:** Resources, Writing – review & editing. **Hayden K.-H. So:** Supervision. **Edmund Y. Lam:** Supervision, Funding acquisition, Writing – review & editing.

Acknowledgement

The work is supported in part by the Research Grants Council of Hong Kong (GRF 17200019, GRF 17201620, GRF 17200321, ECS 27202919), and ACCESS — AI Chip Center for Emerging Smart Systems, sponsored by InnoHK funding, Hong Kong SAR.

References

[1] Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. Annu Rev Pathol 2013;8:331–59.
[2] Liao J, Bian L, Bian Z, Zhang Z, Patel C, Hoshino K, et al. Single-frame rapid autofocusing for brightfield and fluorescence whole slide imaging. Biomed Opt Express 2016;7(11):4763–8.
[3] Hillman EM, Voleti V, Li W, Yu H. Light-sheet microscopy in neuroscience. Annu Rev Neurosci 2019;42:295–313.
[4] Usaj MM, Styles EB, Verster AJ, Friesen H, Boone C, Andrews BJ. High-content screening for quantitative cell biology. Trends Cell Biol 2016;26(8):598–611.
[5] Liron Y, Paran Y, Zatorsky N, Geiger B, Kam Z. Laser autofocusing system for high--resolution cell biological imaging. J Microsc 2006;221(2):145–51.
[6] Guo K, Liao J, Bian Z, Heng X, Zheng G. InstantScope: a low-cost whole slide imaging system with instant focal plane detection. Biomed Opt Express 2015;6(9):3210–16.
[7] Liao J, Jiang Y, Bian Z, Mahrou B, Nambiar A, Magsam AW, et al. Rapid focus map surveying for whole slide imaging with continuous sample motion. Opt Lett 2017;42(17):3379–82.
[8] Ren Z, Chen N, Lam EY. Automatic focusing for multisectional objects in digital holography using the structure tensor. Opt Lett 2017;42(9):1720–3.
[9] Ren Z, Lam EY, Zhao J. Acceleration of autofocusing with improved edge extraction using structure tensor and Schatten norm. Opt Express 2020;28(10):14712–28.
[10] Lichtsteiner P, Posch C, Delbruck T. A 128 × 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE J Solid-State Circuits 2008;43(2):566–76.
[11] Gallego G, Delbrck T, Orchard G, Bartolozzi C, Taba B, Censi A, et al. Event-based vision: a survey. IEEE Trans Pattern Anal Mach Intell 2022;44(1):154–80.
[12] Gehrig D, Rüegg M, Gehrig M, Hidalgo-Carrió J, Scaramuzza D. Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction. IEEE Rob Autom Lett 2021;6(2):2822–9.
[13] Xie Z, Chen S, Orchard G. Event-based stereo depth estimation using belief propagation. Front Neurosci 2017;11:535.
[14] Ge Z, Gao Y, So HK-H, Lam EY. Event-based laser speckle correlation for micro motion estimation. Opt Lett 2021;46(16):3885–8.
[15] Ge Z, Zhang P, Gao Y, So HK-H, Lam EY. Lens-free motion analysis via neuromorphic laser speckle imaging. Opt Express 2022;30(2):2206–18.
[16] Ge Z, Meng N, Song L, Lam EY. Dynamic laser speckle analysis using the event sensor. Appl Opt 2021;60(1):172–8.
[17] Lee JH, Delbruck T, Pfeiffer M, Park PK, Shin C-W, Ryu H, et al. Real-time gesture interface based on event-driven processing from stereo silicon retinas. IEEE Trans Neural Netw Learn Syst 2014;25(12):2250–63.
[18] Moeys DP, Corradi F, Li C, Bamford SA, Longinotti L, Voigt FF, et al. A sensitive dynamic and active pixel vision sensor for color or neural imaging applications. IEEE Trans Biomed Circuits Syst 2017;12(1):123–36.
[19] Ni Z, Pacoret C, Benosman R, Ieng S, Régnier S. Asynchronous event-based high speed vision for microparticle tracking. J Microsc 2012;245(3):236–44.
[20] Howell J, Hammarton TC, Altmann Y, Jimenez M. High-speed particle detection and tracking in microfluidic devices using event-based sensing. Lab Chip 2020;20(16):3024–35.
[21] Bian Z, Guo C, Jiang S, Zhu J, Wang R, Song P, et al. Autofocusing technologies for whole slide imaging and automated microscopy. J Biophotonics 2020;13(12):e202000227.
[22] Liu C-S, Jiang S-H. Precise autofocusing microscope with rapid response. Opt Lasers Eng 2015;66:294–300.
[23] Wei J, Hellmuth T. Optical coherence tomography assisted ophthalmologic surgical microscope. US Patent 5,493,109; 1996.
[24] Cable A, Wollenzin J, Johnstone R, Gossage K, Brooker JS, Mills J, Jiang J, Hillmann D. Microscopy system with auto-focus adjustment by low-coherence interferometry. US Patent 9,869,852; 2018.
[25] Virág T, László A, Molnár B, Tagscherer A, Varga VS. Focusing method for the high--speed digitalisation of microscope slides and slide displacing device, focusing optics, and optical rangefinder. US Patent 7,663,078; 2010.
[26] Silvestri L, Muellenbroich MC, Costantini I, Di Giovanna AP, Sacconi L, Pavone FS. RAPID: Real-time image-based autofocus for all wide-field optical microscopy systems. BioRxiv 2017:170555.
[27] Ou H, Wu Y, Lam EY, Wang B-Z. New autofocus and reconstruction method based on a connected domain. Opt Lett 2018;43(9):2201–3.
[28] Ren Z, Xu Z, Lam EY. Learning-based nonparametric autofocusing for digital holography. Optica 2018;5(4):337–44.
[29] Pinkard H, Phillips Z, Babakhani A, Fletcher DA, Waller L. Deep learning for single-shot autofocus microscopy. Optica 2019;6(6):794–7.
[30] Ren Z, Xu Z, Lam EY. Autofocusing in digital holography using deep learning. In: Three-dimensional and multidimensional microscopy: image acquisition and processing, vol. 10499; 2018. p. 104991V.
[31] Luo Y, Huang L, Rivenson Y, Ozcan A. Single-shot autofocusing of microscopy images using deep learning. ACS Photonics 2021;8(2):625–38.
[32] Zou Y, Crandall GJ, Olson A. Real-time focusing in line scan imaging. US Patent 9,841,590; 2007.

[33] Silfies J.S., Lieser E.G., Schwartz S.A., Davidson M.W.. Nikon perfect focus system (PFS). 2022. https://www.microscopyu.com/applications/live-cell-imaging/nikon-perfect-focus-system.
[34] Sun Y, Duthaler S, Nelson BJ. Autofocusing in computer microscopy: selecting the optimal focus algorithm. Microsc Res Tech 2004;65(3):139–49.
[35] Xi L, Guosui L, Ni J. Autofocusing of ISAR images based on entropy minimization. IEEE Trans Aerosp Electron Syst 1999;35(4):1240–52.
[36] Yazdanfar S, Kenny KB, Tasimi K, Corwin AD, Dixon EL, Filkins RJ. Simple and robust image-based autofocusing for digital microscopy. Opt Express 2008;16(12):8670–7.
[37] Redondo R, Cristóbal G, Garcia GB, Deniz O, Salido J, del Milagro Fernandez M, et al. Autofocus evaluation for brightfield microscopy pathology. J Biomed Opt 2012;17(3):036008.
[38] Subbarao M, Tyan J-K. Selecting the optimal focus measure for autofocusing and depth-from-focus. IEEE Trans Pattern Anal Mach Intell 1998;20(8):864–70.
[39] Neil D, Pfeiffer M, Liu S-C. Phased LSTM: Accelerating recurrent network training for long or event-based sequences. In: Proceedings of the 30th international conference on neural information processing systems; 2016. p. 3889–97.
[40] Gallego G, Lund JE, Mueggler E, Rebecq H, Delbruck T, Scaramuzza D. Event-based, 6-DOF camera tracking from photometric depth maps. IEEE Trans Pattern Anal Mach Intell 2017;40(10):2402–12.
[41] Zhu A.Z., Yuan L., Chaney K., Daniilidis K.. EV-FlowNet: self-supervised optical flow estimation for event-based cameras. arXiv preprint arXiv:180206898 2018.
[42] Maqueda AI, Loquercio A, Gallego G, García N, Scaramuzza D. Event-based vision meets deep learning on steering prediction for self-driving cars. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 5419–27.
[43] Goodman JW. Introduction to Fourier optics. 4th ed. Macmillan Learning; 2017.
[44] Brandli C, Berner R, Yang M, Liu S-C, Delbruck T. A 240 × 180 130 dB 3 µs latency global shutter spatiotemporal vision sensor. IEEE J Solid-State Circuits 2014;49(10):2333–41.