



# Acceleration of autofocusing with improved edge extraction using structure tensor and Schatten norm

ZHENBO REN,<sup>1,3</sup>  EDMUND Y. LAM,<sup>2</sup>  AND JIANLIN ZHAO<sup>1,4</sup>

<sup>1</sup>MOE Key Laboratory of Material Physics and Chemistry under Extraordinary Conditions, and Shaanxi Key Laboratory of Optical Information Technology, School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an 710129, China

<sup>2</sup>Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

<sup>3</sup>zbren@nwpu.edu.cn

<sup>4</sup>jlzhao@nwpu.edu.cn

**Abstract:** Determining the optimal focal plane amongst a stack of blurred images in a short response time is a non-trivial task in optical imaging like microscopy and photography. An autofocusing algorithm, or in other words, a focus metric, is key to effectively dealing with such problem. In previous work, we proposed a structure tensor-based autofocusing algorithm for coherent imaging, i.e., digital holography. In this paper, we further extend the realm of this method in more imaging modalities. With an optimized computation scheme of structure tensor, a significant acceleration of about fivefold in computation speed without sacrificing the autofocusing accuracy is achieved by using the Schatten matrix norm instead of the vector norm. Besides, we also demonstrate its edge extraction capability by retrieving the intermediate tensor image. Synthesized and experimental data acquired in various imaging scenarios such as incoherent microscopy and photography are demonstrated to verify the efficacy of this method.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Optical imaging is an essential and significant technique for people to observe and thus understand the real world. For an optical imaging system, such as photography or microscopy, only the object structure captured at the focal plane, or in a broad sense, within the range of depth-of-focus (DOF), is sharp and clear enough for the subsequent analysis. Structures positioning outside the DOF are considered out-of-focus and thus blurry [1,2]. As is well known, an in-focus image is essential and critical for researchers to perform semantic interpretation and clinical judgment in research fields such as computer vision [3], microscopy [4,5] and digital pathology [6]. Therefore, among a series of images with different amounts of blurriness, it is necessary to automatically recognize the in-focus and sharp image by the use of some proper algorithms. Besides, in optical imaging, abundant images are captured, transmitted and stored for post-processing. Especially for living specimens investigation in microscopy and automatic focus in photography, searching for the optimal focal plane in real time is essential and critical. As such, it is quite imperative to accelerate the computation of autofocusing as fast as possible.

In the area of image processing, edges, which are essential structural features, contain abundant connotation information such as steps, directional changes and the shape of a target. Edges of an image can intuitively outline the target and reflect its local characteristics (grayscale, depth, illumination, texture, etc.), enabling observers to understand the scenario in a more straightforward way. As such, extracting the edge of an image is a basic procedure in image processing and analysis for purposes of segmentation, target recognition and regional shape extraction [7].

Essentially, existing autofocusing methods can be categorized into two common groups, active approach and passive approach [8,9]. The former uses a sensing equipment which can emanate an indicating signal, such as an ultrasound wave or an infrared ray, to record the time of flight between the source and the object of interest. The distance is then estimated according to the transmission duration and the speed of the used ray. Or, with the help of computational imaging principle, by involving additional specially designed hardwares, the focal plane can be successfully determined [10]. The second approach is a direct employment of image processing techniques. A stack of images with different focus settings is captured by an imaging system under the same condition (viewpoint, illumination, etc.). With or without pre-processing the captured images based on phase extraction and wavefront propagation [11,12], an autofocusing algorithm is then applied to assess the degree of image sharpness or blurriness of these images one by one. Such focus measure provides a maximum or minimum value, according to the measurement is the image sharpness or blurriness, when the image is best in-focus. Therefore, the well-focused position is attained with respect to the extreme value and the camera parameters can then be adjusted accordingly for future acquisition [13].

However, both focusing groups have their own advantages and disadvantages. Even though the active approach is fast, it requires an extra sensing equipment which is usually bulky and expensive. The ray emanated from the equipment is harmful to biological specimens and induces fluorescent molecules to be photobleaching. In addition, the maximal distance it can measure is restricted due to the distortion and attenuation of light passing through the air. Despite the passive approach relies only on the quality of captured images and thus is not subjected to the restriction, the algorithm is computationally expensive in determining the sharpness or blurriness of an image sequence. In the meantime, for a three-dimensional (3D) scene containing more than one principal sections, when one section is in-focus, others are out-of-focus and act as defocus noise. As such, current methods are unable to detect the individual distances correctly, or output distorted autofocusing results [14]. In recent years, driven by the development of deep learning and big data, learning-based autofocusing methods have been proposed [15–18]. However, although effective, abundant training data is required in order to get a superior performance. Consequently, they are not universal and not suitable for any scenarios.

Generally, an edge is defined as boundary pixels that connect two separate regions with changing amplitude attributes such as different constant luminance and tristimulus values in an image [19,20]. Therefore, the detection operation usually begins with the examination of the local discontinuity at each pixel element. Commonly-used edge detectors base on various differential operators, including the first- and second-order derivatives, and fitting models [21,22]. One subgroup of the first-order derivative is evaluating the gradients generated along two orthogonal directions. An edge is judged present if the gradient of the image exceeds a user-defined threshold value. Another one is utilizing an elaborately-designed set of discrete edge templates with different orientations. By convoluting the image with the template gradient array, the edge angle is finally determined by the direction of the largest gradient. The second-order derivative bases on the Laplacian operator and its derived templates, in which the Laplacian of Gaussian edge detector is mostly used. As for the fitting model method, for an two-dimensional (2D) image, an ideal 2D step function is first constructed in order to determine whether an edge appears. The edge fitting detector, however, requires more computation in comparison with derivative-based ones.

In this paper, to surmount the dependence of an autofocusing algorithm on a specific imaging modality, on the basis of the method constructed with structure tensor and verified in coherent imaging [13], we further expand the realm of this method to more imaging modalities, incoherent microscopy and photography. To reduce the computational requirement which is mainly caused by using the vector norm and the singular value decomposition (SVD) in calculating the focus metric, we propose to utilize the Schatten norm of a matrix to avoid time-consuming

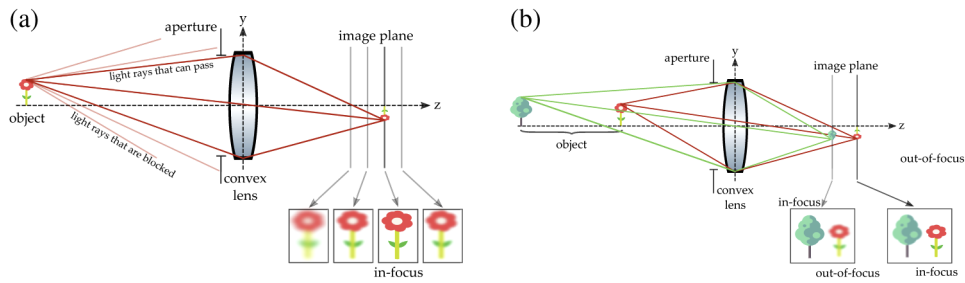
eigendecomposition. Compared with the existing focus metrics, the proposed method not only works for the single-sectional object, but also can handle more complex scenarios, for example, a multi-sectional scene in which multiple focal planes exist in 3D. Besides, we also demonstrate the edge extraction capability of this tensor-based method. The intermediate tensor image can extract the edge information of the individual focal sections with a superior performance over current edge detectors.

## 2. Problem formulation and existing methods

In this section, we introduce the formulation regarding the in-focus and out-of-focus images captured by an imaging system, and present widely-used focus metrics as well as edge detectors for subsequent comparison.

### 2.1. Imaging system

As is known, a linear and time-invariant optical imaging system, such as incoherent microscopy and photography, can be generalized as Fig. 1, which demonstrates the in-focus and out-of-focus geometry using a convex lens. Lights or rays radiating from an object are refracted by the lens and thus converge onto the image plane where a detector is located to capture an image.



**Fig. 1.** (a) Image formation of a single-sectional object. Only on the focal/image plane of the lens, the object in the image is in-focus. (b) Image formation of a multi-sectional object. Even the image is captured on the focal/image plane of the lens, one object in the image is in-focus, while another one is always out-of-focus.

As shown in Fig. 1(a), each point on the object plane is projected onto the image plane and forms one single point. By collecting all the points on the image plane, a well-focused and sharp image is then recorded. When the object or the camera is displaced by a distance, the point originating from the object spreads out and forms a circularly symmetric disk, instead of a single point. The distribution of the pattern leads to an out-of-focus image on the image plane, such that details of the object will be lost and the spatial resolution will be lower. While in Fig. 1(b), a 3D object that contains two discrete 2D objects at different axial positions is to be imaged. Since the focal distance of the lens and the distance of the detector are fixed for a certain acquisition, on the image plane, only one single section is in-focus while another section is inevitably defocused. As such, the resultant image is always blurred [23].

In the mathematical viewpoint, under the coordinate system  $(x, y, z)$ , shortened as  $(\mathbf{x}; z)$ , without considering the noise, the image captured by an imaging system shown in Fig. 1 can be modeled as

$$u(\mathbf{x}) = \sum_{i=1}^N f_i(\mathbf{x}; z) * h_i(\mathbf{x}; z), \quad (1)$$

where  $u(\mathbf{x})$  denotes the captured 2D scalar image,  $N$  is the number of 2D objects to be imaged,  $f_i(\mathbf{x}; z)$  is the 2D (when  $N = 1$ ) or 3D (when  $N \geq 2$ ) unknown and ideal object,  $h_i(\mathbf{x}; z)$  represents

the time-invariant point spread function (PSF) of the imaging system, and  $*$  stands for the 2D convolution [23]. For imaging systems such as microscopy and photography, the PSF  $h_i(\mathbf{x}; z)$  is also space-invariant. Therefore, for a 3D object, if one section is captured clearly, other sections at different longitudinal depths are out-of-focus and manifest as defocus noise.

## 2.2. Existing methods

### 2.2.1. Focus metrics

Numerous focus metrics have been proposed to evaluate the sharpness or blurriness of images. Generally, image-based methods are categorized into four groups: derivative-based, statistics-based, histogram-based and intuitive algorithms [24]. In this paper, we select four typical and widely-used focus metrics from each group, i.e., the  $\ell_1$ -norm of image gradient, variance, entropy and image power, to compare with the proposed focus measure [25–28]. The definition of the first selected focus metric,  $\ell_1$ -norm of image gradient, is given as

$$M_1 = \int \left| u_x \right| + \left| u_y \right| d\mathbf{x}, \quad (2)$$

where  $u_x$  and  $u_y$  denote the gradient  $\nabla u(\mathbf{x}) = (u_x, u_y)^T$  at pixel  $\mathbf{x}$ . The second metric, variance, is defined as

$$M_2 = \int \left( u(\mathbf{x}) - \mu \right)^2 d\mathbf{x}, \quad (3)$$

where  $\mu$  is the mean intensity of  $u(\mathbf{x})$ . The third one is entropy and it reads

$$M_3 = \int p_i \log_2(p_i) d\mathbf{x}, \quad (4)$$

where  $p_i$  is the probability of a pixel  $\mathbf{x}$  with intensity  $i$ . The last one is image power, which is given as

$$M_4 = \int \left( u(\mathbf{x}) \right)^2 d\mathbf{x}. \quad (5)$$

To quantitatively evaluate the performance of the individual autofocus algorithms, we take the *unimodality*, *accuracy*, *reproducibility* and *range* [29,30] as criteria. The unimodality ensures that only one sole extremum exists at the focal position within the whole searching range. When there are multiple focal planes, the unimodality still remains true with multiple extrema, correspondingly. The accuracy measures the error between the real focus position and the estimated position using a focus measure. The third one, reproducibility, is calculated as the width at a high percentage, normally 80%, of the maximal value. A sharp top of the extremum results in a good reproducibility. Last but not the least, the range describes how long the extremum is tailed, and is computed as the full width at half maximum (FWHM). In the following comparison, autofocus values along the vertical axis are normalized within the range [0, 1] for a better visualization.

### 2.2.2. Edge extraction

As discussed above, edge extraction is a classical problem in image processing and thus numerous methods have been demonstrated. In this paper, similarly, we select four widely used edge detectors, Sobel, Prewitt, Roberts and Canny, for qualitative and quantitative comparison with the proposed structure tensor. Since they are well-established techniques, we refer readers to Ref. [22] for the individual definitions and implementations.

For the performance evaluation of edge extraction, we use Pratt's Figure of Merit (FoM) [31] as a metric to examine edge extractors. FoM generates maximum performance for well localized

and continuous edges, and penalizes for missing edge points and noise classified as edge pixels. It is defined as

$$\text{FoM} = \frac{1}{\max(I_I, I_A)} \sum_{i=1}^{I_A} \frac{1}{1 + \alpha d^2}, \quad (6)$$

where  $I_I$  and  $I_A$  represent the number of ideal and actual edge map points,  $\alpha$  is a scaling constant and  $d$  is the separation distance of an actual edge point normal to a line of ideal edge points. The rating factor  $\text{FoM} = 1$  means that the edge is precisely detected, while  $\text{FoM} = 0$  means the worst detection. The value of scaling factor,  $\alpha$ , reflects the performance on edge localization and is set as  $1/9$  by experience. A smeared edge is easier to be fixed than an offset edge since a smeared edge can be thinned by morphological post-processing operation.

### 3. Acceleration and edge-preserving autofocusing

#### 3.1. Formation of vector-norm-based autofocusing method

In communities of image processing and computer vision, structure tensor, also referred to as the second-moment matrix, is a widely used tool for problems like corner detection, interest point detection, and feature tracking. It is a matrix derived from the gradient of a function, or more specifically, of an image, and summarizes the predominant directions of the gradient in a specified neighborhood of a point, and the degree to which those directions are coherent [32,33].

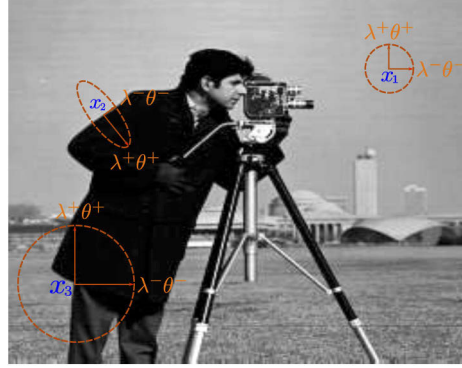
For an image  $u(\mathbf{x})$ , the 2D structure tensor  $\mathbf{S}$  at pixel element  $\mathbf{x}$  can be written as

$$\begin{aligned} \mathbf{S}(\mathbf{x}) &= G(\mathbf{x}) * [\nabla u(\mathbf{x}) \nabla u(\mathbf{x})^T] \\ &= G(\mathbf{x}) * \begin{bmatrix} \left(\frac{\partial u(\mathbf{x})}{\partial x}\right)^2 & \left(\frac{\partial u(\mathbf{x})}{\partial x} \frac{\partial u(\mathbf{x})}{\partial y}\right) \\ \left(\frac{\partial u(\mathbf{x})}{\partial y} \frac{\partial u(\mathbf{x})}{\partial x}\right) & \left(\frac{\partial u(\mathbf{x})}{\partial y}\right)^2 \end{bmatrix}, \end{aligned} \quad (7)$$

where  $\nabla u(\mathbf{x})$  is the 2D spatial intensity gradient,  $[\cdot]^T$  denotes matrix transpose,  $G(\mathbf{x})$  is a 2D Gaussian function that performs smoothing operation within a window [34,35]. The 2D structure tensor  $\mathbf{S}(\mathbf{x}) \in \mathbb{S}_+^2$  of an image at an image point  $\mathbf{x}$  is a symmetric and semi-positive-definite matrix, such that it has two positive eigenvalues, which lead to the importance of this concept. Let  $\lambda^+ = \Lambda^+(\mathbf{S}(\mathbf{x}))$ ,  $\lambda^- = \Lambda^-(\mathbf{S}(\mathbf{x}))$ , where  $\Lambda(\cdot)$  denotes an operator getting the eigenvalues, and  $\theta^+$ ,  $\theta^-$  be the corresponding unit eigenvectors, thus we have  $\lambda^+ \geq \lambda^- \geq 0$ . The structure tensor measures the geometry of image structures in the neighborhood of each point. Its eigenvectors  $\theta^+$  and  $\theta^-$  describe the orientation of maximum and minimum vectorial variation of  $u(\mathbf{x})$ , and its eigenvalues  $\lambda^+$  and  $\lambda^-$  describe measures of these variations. Therefore the eigenvalues of the structure tensor offer a rich and discriminative description of the local geometry of the image [13].

More precisely, when both  $\lambda^+$  and  $\lambda^-$  are relatively small, as the point  $\mathbf{x}_1$  in Fig. 2 shows, the variation around the point is small. This indicates that the region is homogeneous and flat, and no change happens in all directions. When  $\lambda^+$  is large and  $\lambda^-$  is small, as the point  $\mathbf{x}_2$  in Fig. 2 shows, there are strong variations, but only in a dominant direction. There is no change along the edge direction. Therefore, point  $\mathbf{x}_2$  is located close to an edge, and its corresponding eigenvector  $\theta^+$  directs to the direction with the most significant change. When both  $\lambda^+$  and  $\lambda^-$  are large, as the point  $\mathbf{x}_3$  in Fig. 2 shows, there are strong variations in all directions, implying that the current point is close to an corner.

Based on the principle explained above, for a specific image  $u(\mathbf{x}) \in \Omega$  in a stack of  $N$  images  $u_n(\mathbf{x})$ ,  $n = 1, 2, \dots, N$ , at each pixel  $\mathbf{x}$  we compute the structure tensor  $\mathbf{S}(\mathbf{x})$  within a window



**Fig. 2.** Sample points  $x_1, x_2, x_3$  denote the flat region, edge and corner, respectively. The structure tensor at a point is visualized as an ellipse and its unit eigenvectors  $\theta^+, \theta^-$  and eigenvalues  $\lambda^+, \lambda^-$  are depicted.

function  $G(\mathbf{x})$ , the size of which is defined accordingly. The focus metric can be computed as

$$M_p(u) = \int_{\Omega} \left\| \lambda \right\|_p \, d\mathbf{x}, \tag{8}$$

where  $\left\| \lambda \right\|_p = \left( \sum_{i=1}^2 \lambda_i^p \right)^{1/p}$ .

### 3.2. Formation of matrix-norm-based autofocusing method

In the previous proposal, eigenvalues in Eq. (8) are calculated using SVD, which requires significant calculation and thus time-consuming. To accelerate the computation, an alternative definition of the focus metric is formulated as follows.

Equivalently, instead of using the vector norm, the matrix norm, or in other words, Schatten norm, is utilized to define the focus metric here [36]. In general, suppose a matrix  $\mathbf{S} \in \mathbb{C}^{N_1 \times N_2}$  has the SVD  $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\dagger$ , where  $\mathbf{U}$  and  $\mathbf{V}$  are  $N_1 \times N_1$  and  $N_2 \times N_2$  unitary matrices consisting of the singular vectors of  $\mathbf{S}$ ,  $\mathbf{\Lambda}$  is an  $N_1 \times N_2$  diagonal matrix consisting of the singular values of  $\mathbf{S}$  respectively, and  $\dagger$  denotes the conjugate transpose. The Schatten norm of order  $p$  of matrix  $\mathbf{S}$  is thus defined as

$$\begin{aligned} \left\| \mathbf{S} \right\|_p &= \left( \sum_{i=1}^{\min(N_1, N_2)} \lambda_i^p(\mathbf{x}) \right)^{1/p} \\ &= \left[ \text{Tr} \left( \sqrt{(\mathbf{S}^\dagger \mathbf{S})^p} \right) \right]^{1/p} \\ &= [\text{Tr}(\mathbf{S}^p)]^{1/p}, \end{aligned} \tag{9}$$

where  $\lambda_i$  is the  $i$ -th singular value of  $\mathbf{S}$ , which corresponds to the  $(i, i)$  entry of  $\mathbf{\Lambda}$ ,  $\mathbf{S}^\dagger = \mathbf{S}$  since the structure tensor matrix  $\mathbf{S}$  is real and symmetric, and  $\text{Tr}(\cdot)$  is to acquire the trace of a matrix. As such, an alternative formulation of the tensor-based autofocusing method can be also defined as

$$M'_p(u) = \int_{\Omega} \left\| \mathbf{S} \right\|_p \, d\mathbf{x}. \tag{10}$$

The Schatten matrix norm of order  $p$  corresponds to the  $\ell_p$ -norm of the vector-norm of the singular values of the matrix. Norms of an image measure the image variation more coherently and

robustly than the gradient magnitude, as they take into account the variations in its neighborhood. In the meanwhile, they incorporate richer information, since they depend both on the maximum and minimum of the directional variation. In this way, this kind of representation is in general better adapted to the image geometry. Therefore, provided we can obtain the  $p$  power of  $\mathcal{S}$ , the focus metric can be easily computed without implementing the time-consuming SVD. In the following computation,  $p$  is set to 3. Note that according to the massive trial and error experiments, although this hyper-parameter changes the calculated sharpness  $M'_p$  of an image  $u(\mathbf{x})$ , it does not affect the correctness of this method, meaning that the position where an extremum occurs does not change.

We then theoretically analyze the acceleration by examining the computational complexity of using matrix norm and vector norm. Since in Ref. [13], the eigenvalues of each matrix through SVD and the  $\ell_p$ -norm computation of the eigenvalues vector have to be computed. For an  $N \times N$  image, the computation of the structure tensor pixel-by-pixel requires  $O(2N^2 \log(N))$  operations, while the computational complexity of SVD decomposition is of order  $O(N^3)$  [37]. Although conventional autofocusing algorithms include the elementwise operations, the subsequent computation is normally of order  $O(N^2)$ . As such, the structure tensor-based method leads to computational demanding compared to conventional autofocusing algorithms. In this paper, we use Schatten matrix norm given in Eqs. (9) and (10), instead of the vector norm of eigenvalues, to compute the focus metric. Since the structure tensor matrix  $\mathcal{S}$  is a real and symmetric matrix, it is diagonalizable as a result. Therefore, taking the  $p$  power can be done in time of  $O(D(N) + N \log(N))$ , where  $D(N)$  is the time to diagonalize a matrix and can be done in  $O(\log(N))$  by using the parallel Jacobi algorithm.

## 4. Results

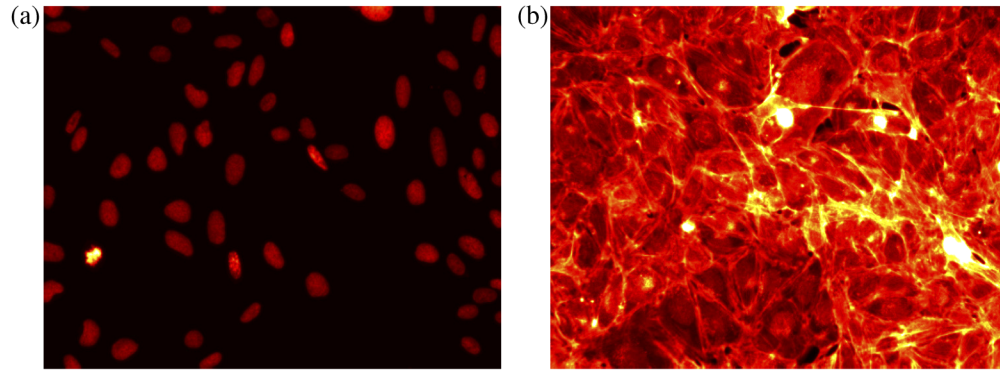
As discussed above, the proposed focus metric is a measure of image sharpness and can be used to find the optimal focal plane in the problem of autofocusing. Therefore, this is a versatile tool and independent of imaging systems. In the following sections, we will proceed to demonstrate its capability in detecting (multiple) focal planes in incoherent microscopy and photography.

### 4.1. Microscopy

The first application is to find the focal plane among a stack of microscopy images. Here we use the publicly available collection of microscopy image database BBBC0006v1 in the Broad Bioimage Benchmark Collection [38,39]. The high-content screening microscopy images were acquired using an ImageXpress Micro automated cellular imaging system (Molecular Devices, Sunnyvale, CA). The database contains 52224 images totally, recording 384-well plates containing human U2OS cells, each of which were stained with Hoechst 33342 and Alexa Fluor 594 Phalloidin markers imaged with an exposure of 15 ms and 1000 ms for Hoechst and Phalloidin, respectively, at 20 $\times$  magnification and 2 $\times$  binning, and 2 sites per well. Microscopic specimens are static and images are captured in the fluorescent modality and a high signal-to-noise ratio (SNR) is configured for a better observation. For a specific image stack, there are 34 images with 2  $\mu$ m between slices altogether, consisting of 696  $\times$  520 pixels in 16-bit TIF format. For each site, the optimal focus was found using laser autofocusing to find the well bottom. The automated microscope was then programmed to collect a z-stack of 32 image sets. The in-focus image should lie in between position 11 and position 23 [5].

Here we select two stacks of images of different cells, Hoechst and Phalloidin, as examples for microscopic imaging system. Focal images of them are presented in Fig. 3, and the two stacks of images are shown in Figs. 4(a) and 4(b). The Hoechst images are simpler and less complex than the Phalloidin images. For each stack the 34 images are aligned in a 6  $\times$  6 grid. From left to right and from top to bottom, the microscopy image gradually turns sharp, and finally turns blurry again. With our naked-eyes we can only tell which images are sharper, but it is difficult to

assert at which position the image is best focused. Thanks to the autofocusing algorithms, we are able to find out the optimal in-focus image with the sharpest and clearest texture for further investigation like cytometry and cell segmentation.



**Fig. 3.** Focal images of (a) Hoechst at position 19 and (b) Phalloidin at position 20.

The focus measure results are shown in Figs. 4(c) and 4(d). According to the proposed method, for Hoechst cell, the image at position 19 is in-focus, while for Phalloidin cell, position 20 is the focal plane. The results are coincident to the information that the in-focus lies in plane 11 – 23, given by the providers of the dataset. Although the five focus metrics have slightly different candidates, they are still credible, except for the entropy of Phalloidin cell. Due to the complexity and low variance of intensities in the sample, it has a wrong detection. Figures 4(e) and 4(f) give the respective edge detections.

Tables 1 and 2 give the quantitative comparison results. We normalize the horizontal axis to [0, 1] and compute the width. Since the actual and focal positions of the specimen are not provided, the accuracy is denoted as “N/A” (not available). As can be seen in the tables, the proposed autofocusing algorithm is unimodal and outperforms the rest in reproducibility and range. The range of variance is “NaN”, since at the right-hand side of the maximum, the curve does not reach half of the maximal value. These two tables demonstrate the superiority of the proposed algorithm.

**Table 1. Comparison of five autofocusing algorithms for Hoechst images**

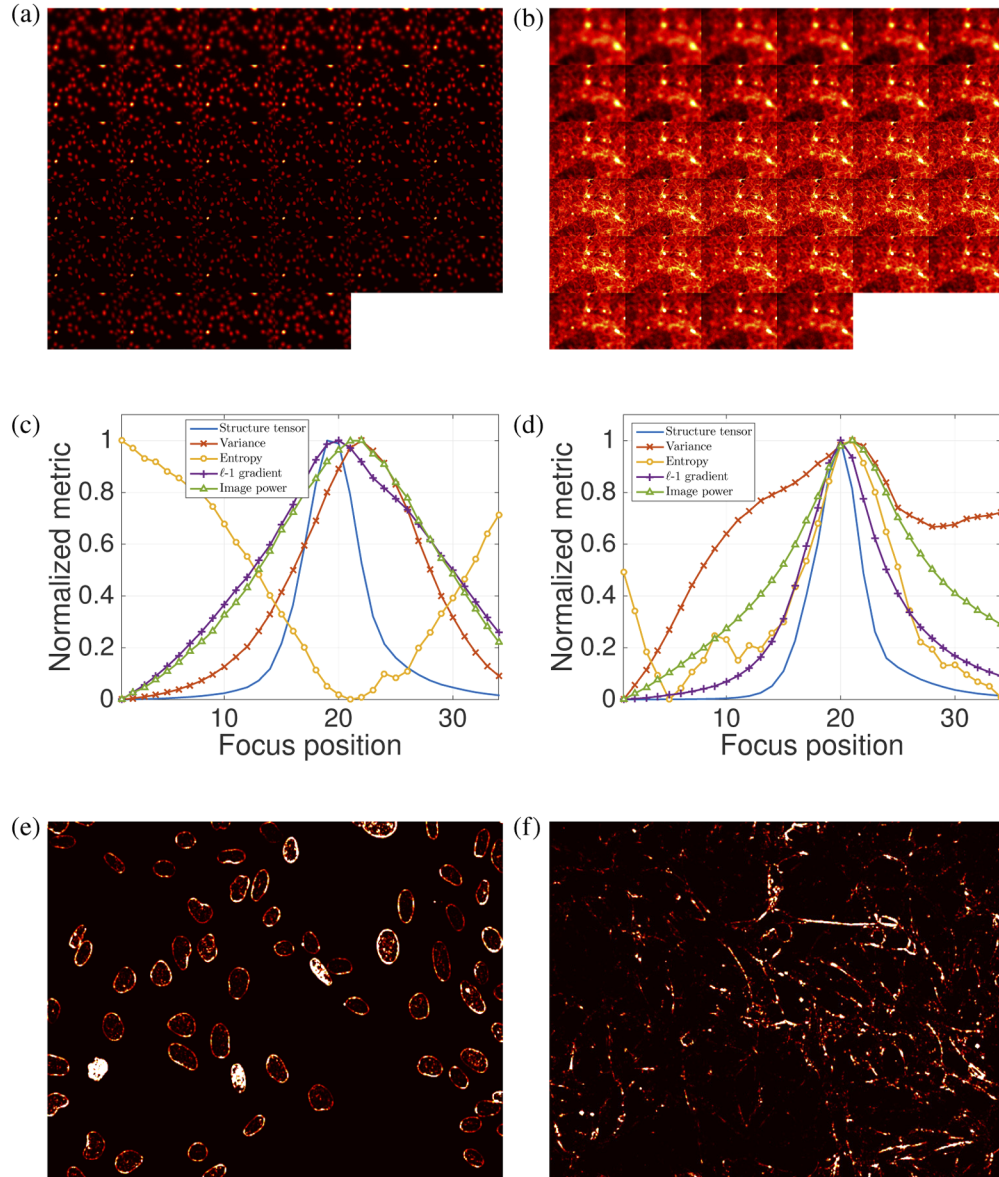
Algorithm	Unimodality	Accuracy	Reproducibility	Range
(fast) Structure tensor	Yes	N/A	<b>0.0622</b>	<b>0.1120</b>
Variance	Yes	N/A	0.1382	0.2536
Entropy	No	N/A	0.3030	0.3818
$\ell_1$ gradient	Yes	N/A	0.1580	0.3591
Image power	Yes	N/A	0.1836	0.3484

To quantitatively verify the acceleration of the proposed fast method, we record the computation time and the original SVD-based method by averaging the autofocusing time in Fig. 4. The former consumes 2.438 s (about 14 frames per second@696 × 520), while the latter takes 12.267 s (about 2.8 frames per second@696 × 520), on a PC with a CPU of Xeon W-2145@3.70GHz with 8 cores. Without losing the accuracy of autofocusing, by using the Schatten matrix norm, the proposed method can indeed speed up the whole calculation, in which parallel computing is even not utilized. Such acceleration, therefore, can facilitate the employment of the proposed method in real-time monitoring living specimens in microscopy.



Table 2. Comparison of five autofocusing algorithms for Phalloidin images

Algorithm	Unimodality	Accuracy	Reproducibility	Range
(fast) Structure tensor	Yes	N/A	<b>0.0600</b>	<b>0.1179</b>
Variance	Yes	N/A	0.3014	NaN
Entropy	No	N/A	0.1699	0.2551
$\ell_1$ gradient	Yes	N/A	0.1058	0.2311
Image power	Yes	N/A	0.1770	0.4066



**Fig. 4.** (a-b) Image sequences of the Hoechst and Phalloidin. (c-d) Focus measures of (a) and (b). (e-f) Edge-preserving images of Hoechst and Phalloidin at the respective focal plane.

## 4.2. Photography

The second application is to determine the sharpness and focal plane in photography. In this subsection, two examples, image processing of Gaussian and motion blur and two-sectional imaging, are demonstrated.

### 4.2.1. Gaussian and motion blur

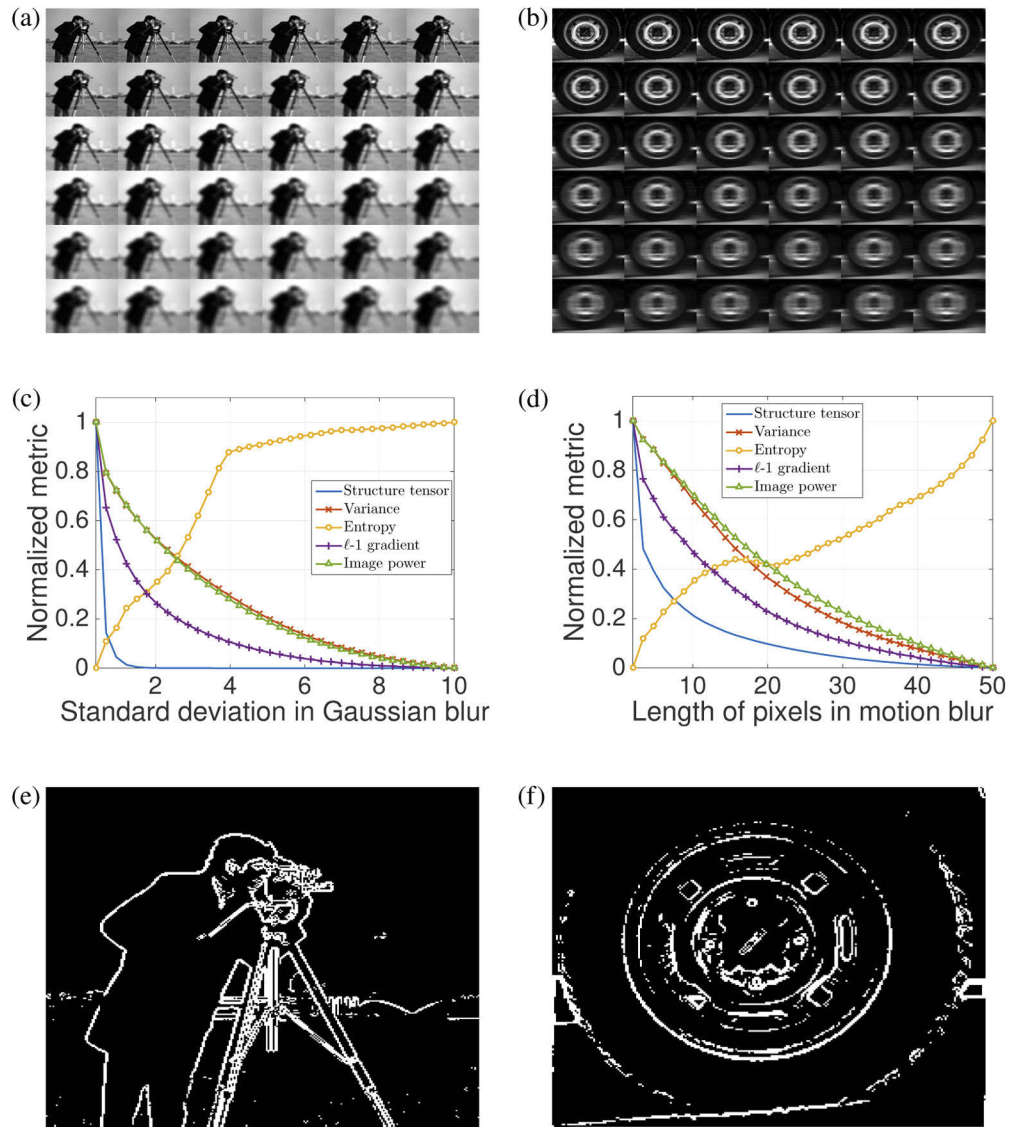
Gaussian blur and motion blur are two commonly seen effects in image processing and photography. As is known, a blurred image, no matter the blur is due to the Gaussian blur [40,41] or motion blur [42], is less sharper. Among a stack of images, the sharpest image is always preferred and needs to be found out. Thus we simulate the two kinds of blur and use focus metrics to determine the best in-focus image. For the Gaussian blur, by increasing the standard deviation, the resultant image becomes more blurry. While for the motion blur, we increase the length of pixels to introduce more blur.

In Figs. 5(a) and 5(b), from top to bottom and from left to right, images are posed more and more blurriness due to the increasing of Gaussian blur and motion blur. We then utilize the focus metrics for autofocusing, and the focus curves for the two cases are shown in Figs. 5(c) and 5(d). Note that the image at position 1 contains no blur and thus has the best quality compared to blurred images. From Figs. 5(c) and 5(d) we can see that, as the amount of blurriness increases, focus scores obtained with different methods decrease as expected. Although the five focus metrics present competitive results visually, the tensor-based method drops to zero quickly. This means that for the *unimodality* and *accuracy*, they have similar performance. The proposed method, however, wins in the *reproducibility* and *range*. Detected edge images of the focal ones, i.e., the original images, in Figs. 5(a) and 5(b) are demonstrated in Figs. 5(e) and 5(f), respectively. As for the time consumption of the matrix-norm method and vector-norm method using structure tensor, images here have a size of  $512 \times 512$ , such that the fast method needs 2.066 s, while the original method takes 13.581 s.

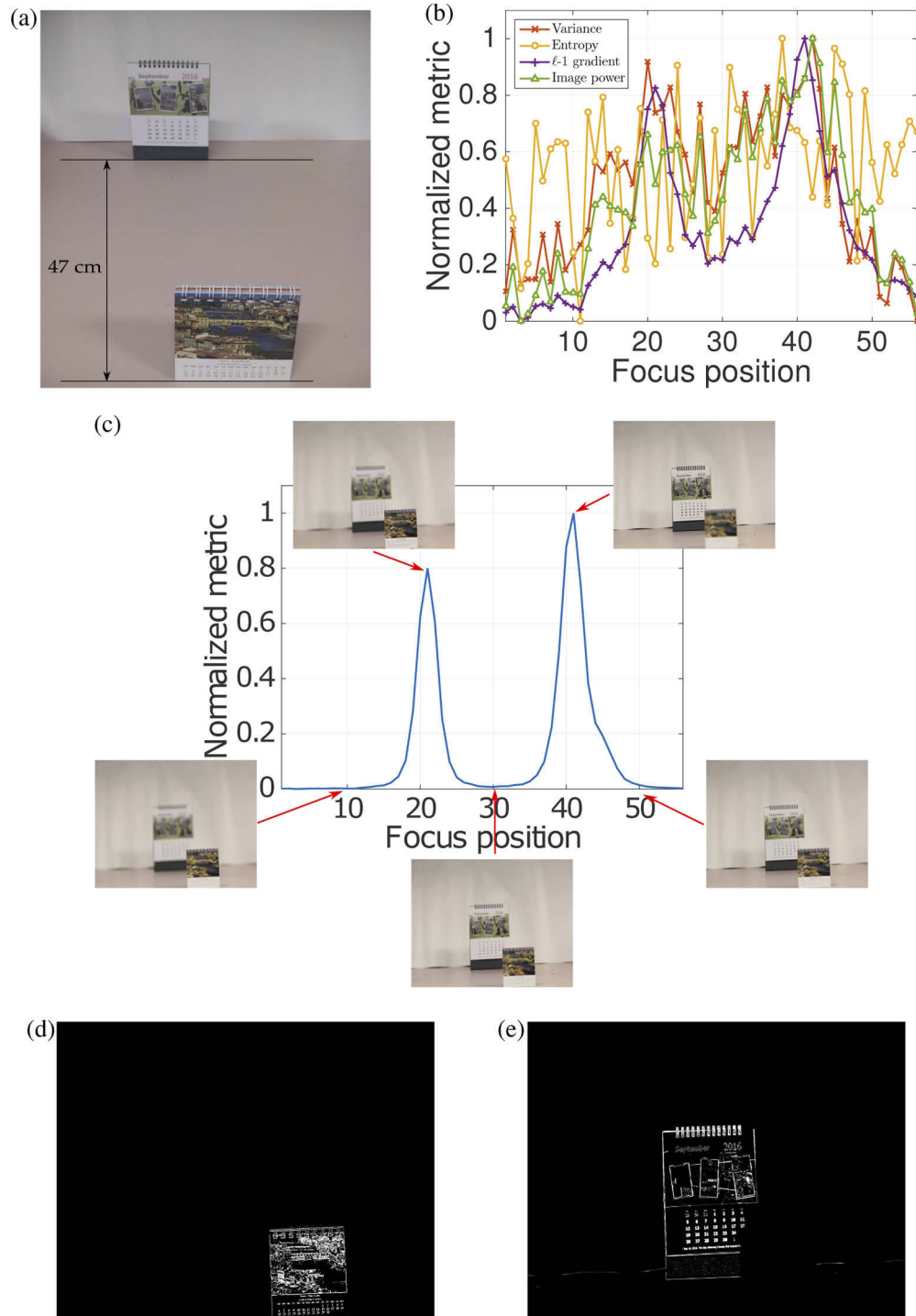
### 4.2.2. Two-sectional imaging

To verify the capability in searching optimal focal planes of a 3D scene, in Fig. 6(a), we construct an object scene for image acquisition. Two discrete calendars, which are regarded as two sections, are located at different longitudinal positions, having a slight overlapping region. A commercial camera (Canon EOS 600D) was used to take totally 56 images under identical illumination and setting (Exposure mode: Manual; Time: 1/40 second; Aperture:  $f\# = 3.5$ ; ISO: 400) from the nearest point (in front of the first calendar) to the farthest point (on the back wall). The distance between the two objects is around 47 cm, and there are 20 images in between. The autofocusing result and selected images are shown in Figs. 6(b) and 6(c). For a clear visualization, we separate the autofocusing result of the proposed method from other four metrics.

As the focus of the camera moves from the nearest position to the farthest position, the blurriness of the two objects is also changing. In Fig. 6(b), only the  $\ell_1$  gradient outputs two maximal points. However, the ripples in the curve bring artifacts and fake extreme values, and make it difficult to formulate the autofocusing as an optimization problem for automatic searching. While in Fig. 6(c), the curve is smooth and two maximal points are obviously manifested at position 21 and position 41. The two maximums represent the two focal planes where the two discrete sections are located. Such comparison clearly shows the superiority of the proposed method. Besides, the extracted edge images in Figs. 6(d) and 6(e) give additional capability of this method. Due to the overlapping, in Fig. 6(e) the lower right corner of the rear calendar is missing, and the white line denotes the edge of the board and the table. Because of the pseudo-random results and failure of other four focus metrics, only the proposed method can successfully find the focal images, leading to the unavailability of quantitative comparison. However, from Fig. 6(c) we can clearly see that the proposed method satisfies the *unimodality* and *accuracy*, and has



**Fig. 5.** (a) Images blurred by a 2D Gaussian function of different standard deviations. (b) Images added with motion blur of different lengths of pixels. (c-d) Focus measurements of Gaussian blur and motion blur. (e-f) Edge detection images of Gaussian blur and motion blur.



**Fig. 6.** (a) Two-sectional object used in the experiment. (b) Autofocusing results using other four focus metrics. (c) Autofocusing result with the proposed method, and images with different blurriness are shown at each position (for full size images, we refer readers to <https://github.com/thomas0708/pictures2020>). (d-e) Extracted edge images.

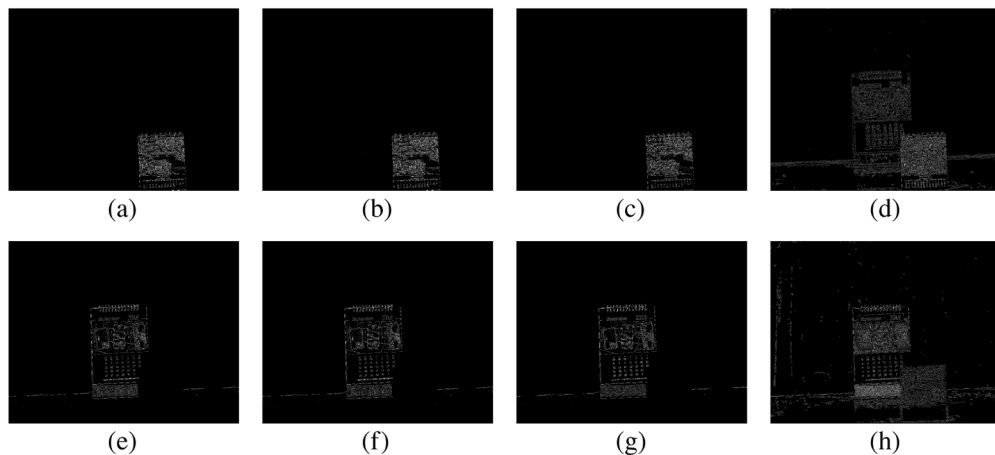
superior performance in *reproducibility* (sharp top) and *range* (short tail). We also record the computation speed of the proposed Schatten norm method and the original method to verify its acceleration. For 56 images with a size of  $5184 \times 3456$  within the stack, the fast method consumes 43.769 s totally, and the original method takes 232.395 s. The improvement in time consumption is about fivefold, indicating a significant acceleration in computation speed.

## 5. Discussions

In this section, we analyze the edge detection performance, characteristics, contrast invariance and computational complexity, of the proposed algorithm.

### 5.1. Edge extraction comparison

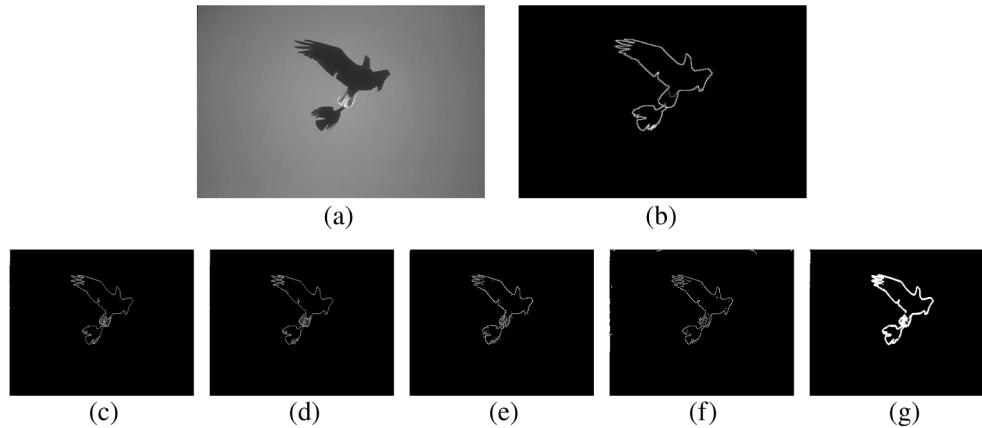
In the above sections, we have shown that the structure tensor-based autofocusing method is capable of extracting the edge information of an image. To further examine the performance of its edge detection capability, we select four widely used edge detectors for qualitative and quantitative comparison. In Fig. 7, edge images of the two-sectional objects obtained with the four methods are demonstrated. Compared with Figs. 6(d) and 6(e), for a multi-sectional object in which one section is in-focus while another one is out-of-focus and blur, except the Canny extractor, other three methods and the proposed one can generate reasonable edge images.



**Fig. 7.** Extracted edge images of the two individual focal images at position 21 and 41 in Fig. 6(a) using conventional detectors of (a) and (e) Sobel, (b) and (f) Prewitt, (c) and (g) Roberts and (d) and (h) Canny.

Furthermore, to quantitatively evaluate the performance, we select an image, as shown in Fig. 8(a), from the Berkeley Segmentation Data Set (BSDS300) [43]. Since its corresponding human-labeled contour map is available and given in Fig. 8(b), therefore the FoM can be calculated for a quantitative comparison. Edge images acquired with the individual extractors are shown in Figs. 8(c)–8(g), and the respective FoM values are given in Table 3. From an illustrative viewpoint, for a single-sectional and in-focus image, all the five extractors can give a reasonable result. However, from the FoM values we know that, the structure tensor-based method performs the best and is thus superior to other four methods. Besides, to quantify the time consumption of the accelerated method and the original one in detecting edges, respective running time on Fig. 8(a) is tested. The image size is  $481 \times 321$ , and the fast method consumes 86.1 ms, while the original one takes 124.2 ms. Acceleration in edge detection is therefore verified. Although other

four conventional edge extractors have a relatively less running time (all are around 50 ms), their FoM results are greatly lower than the proposed method.



**Fig. 8.** (a) A natural image and (b) ground-truth contour map. Extracted edge images using detectors of (c) Sobel, (d) Prewitt, (e) Roberts, (f) Canny and (g) structure tensor.

**Table 3. Comparison of FoM values among five edge detectors in Fig. 8.**

Method	(fast) Structure tensor	Sobel	Prewitt	Roberts	Canny
FoM	<b>0.6602</b>	0.1939	0.1944	0.2393	0.2133

### 5.2. Monotonicity and unimodality

A basic requirement for a focus metric is that within a searching range or among a sequence of images, the extremum is reached only at the focal plane. For a single-sectional scenario, out-of-focus images should have a lower/higher (depending on the characteristic of the focus metric) score than the in-focus image. Such feature leads to the so-called monotonicity and unimodality. This is also true for a multi-sectional scene where monotonicity holds before and after an extremum, and unimodality guarantees one focal plane corresponds one extremum. Multiple ripples or fluctuations lead to a failure of correctly recognizing focal planes.

We have shown that for the case of  $p = 1$ , when a larger degree of blurring occurs, a smaller value of  $M_p(u)$  will be obtained [13]. Consequently, such argument ensures demands of monotonicity and unimodality. For a higher order of  $p$ , according to the Minkowski inequality, the  $\ell_p$ -norm in the  $L^p$  space is convex for any  $p \geq 1$ . Additionally, we know that the integral of the  $\ell_p$ -norm does not change its convexity. As a consequence, the structure tensor-based focus metric is also convex [44]. This property results in the unimodality within the range of in-focus and out-of-focus positions. As such, the monotonicity is also guaranteed when the blurriness increases or decreases monotonically.

### 5.3. Contrast invariance

Invariance requirements play a central role in image analysis because the objects must be recognized under varying conditions of illumination (contrast invariance) and from different points of view (projective invariance). Although in this paper, we are dealing with images acquired at one fixed point of view and thus fixed illumination. Therefore, for a focus metric, it is necessary to guarantee the invariance of contrast amongst an image sequence. Suppose the contrast level of an  $N \times N$  image  $f(x, y)$  is globally modified by a positive factor  $\alpha$ , we then have

a new image  $g(x, y) = \alpha f(x, y)$ . Then the contrast invariance can be achieved by normalizing the new image  $g(x, y)$  with its energy as

$$\begin{aligned}
 \tilde{g}(x, y) &= \frac{g(x, y)}{\sqrt{\sum_{x=1}^N \sum_{y=1}^N [g(x, y)]^2}} \\
 &= \frac{\alpha f(x, y)}{\sqrt{\sum_{x=1}^N \sum_{y=1}^N [\alpha f(x, y)]^2}} \\
 &= \frac{f(x, y)}{\sqrt{\sum_{x=1}^N \sum_{y=1}^N [f(x, y)]^2}} \\
 &= \tilde{f}(x, y).
 \end{aligned} \tag{11}$$

As such, the focus metric derived from the normalized stack of images is guaranteed to be contrast invariant. This pre-processing step is normally included before the computation of focus metric, and is crucial particularly for images captured under different illumination conditions to provide more consistent results.

## 6. Conclusion

In this paper, we propose an optimization strategy to accelerate the computation of structure tensor for autofocusing. Main contributions lie in three aspects: (1) By using the Schatten matrix norm instead of the vector norm, time-consuming eigenvalue decomposition is avoided and the computation speed can be significantly improved by about fivefold. (2) The efficacy of the tensor-based autofocusing method is verified in two typical incoherent imaging modalities, the wide-field microscopy and photography. The realm of this method is expanded. (3) Compared to conventional autofocusing methods, an additional capability of the proposed method in edge extraction by using the intermediate tensor image is demonstrated. Qualitative and quantitative comparison demonstrate that this method outperforms the existing focus metrics in autofocusing, especially for the scenario where multiple focal planes exist. This method also gives the best edge extraction results compared with the widely used edge detectors.

## Funding

National Natural Science Foundation of China (61905197); Fundamental Research Funds for the Central Universities (310201911qd002).

## Acknowledgments

We used the image set BBBC006v1 from the Broad Bioimage Benchmark Collection [39].

## Disclosures

The authors declare no conflicts of interest.

## References

1. Z. Ren, N. Chen, and E. Y. Lam, "Extended focused imaging and depth map reconstruction in optical scanning holography," *Appl. Opt.* **55**(5), 1040–1047 (2016).
2. Z. Ren, Z. Xu, and E. Y. Lam, "End-to-end deep learning framework for digital holographic reconstruction," *Adv. Photonics* **1**(1), 1 (2019).
3. E. Krotkov, "Focusing," *Int. J. Comput. Vision* **1**(3), 223–237 (1988).
4. X. Zhang, E. Y. Lam, T. Kim, Y. S. Kim, and T.-C. Poon, "Blind sectional image reconstruction for optical scanning holography," *Opt. Lett.* **34**(20), 3098–3100 (2009).

5. R. Lenz, "Generalized Pareto distributions-Application to autofocus in automated microscopy," *IEEE J. Sel. Top. Signal Process.* **10**(1), 92–98 (2016).
6. A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Med. Image Anal.* **33**, 170–175 (2016).
7. R. Maini and H. Aggarwal, "Study and comparison of various image edge detection techniques," *Int. J. Image Process.* **3**, 1–11 (2008).
8. S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.* **46**(5), 1415–1432 (2013).
9. C. Y. Wee and R. Paramesran, "Measure of image sharpness using eigenvalues," *Inf. Sci.* **177**(12), 2533–2552 (2007).
10. J. Liao, L. Bian, Z. Bian, Z. Zhang, C. Patel, K. Hoshino, Y. C. Eldar, and G. Zheng, "Single-frame rapid autofocusing for brightfield and fluorescence whole slide imaging," *Biomed. Opt. Express* **7**(11), 4763–4768 (2016).
11. J. Xu, X. Tian, X. Meng, Y. Kong, S. Gao, H. Cui, F. Liu, L. Xue, C. Liu, and S. Wang, "Wavefront-sensing-based autofocusing in microscopy," *J. Biomed. Opt.* **22**(08), 1 (2017).
12. J. Xu, Y. Kong, Z. Jiang, S. Gao, L. Xue, F. Liu, C. Liu, and S. Wang, "Accelerating wavefront-sensing-based autofocusing using pixel reduction in spatial and frequency domains," *Appl. Opt.* **58**(11), 3003–3012 (2019).
13. Z. Ren, N. Chen, and E. Y. Lam, "Automatic focusing for multisectional objects in digital holography using the structure tensor," *Opt. Lett.* **42**(9), 1720–1723 (2017).
14. P. T. Yap and P. Raveendran, "Image focus measure based on Chebyshev moments," *IEE Proc., Vis. Image Process.* **151**(2), 128 (2004).
15. Z. Ren, Z. Xu, and E. Y. Lam, "Autofocusing in digital holography using deep learning," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXV*, vol. 10499 (International Society for Optics and Photonics, 2018), p. 104991V.
16. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica* **5**(4), 337–344 (2018).
17. S. Jiang, J. Liao, Z. Bian, K. Guo, Y. Zhang, and G. Zheng, "Transform- and multi-domain deep learning for single-frame rapid autofocusing in whole slide imaging," *Biomed. Opt. Express* **9**(4), 1601–1612 (2018).
18. H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, "Deep learning for single-shot autofocus microscopy," *Optica* **6**(6), 794–797 (2019).
19. T.-H. H. Lee, "Edge detection analysis," *Int. J. Comput. Sci. Issues* **5**, 1–25 (2007).
20. Z. Ren and E. Y. Lam, "Edge-preserving autofocusing in digital holography," in *Digital Holography and Three-Dimensional Imaging*, (Optical Society of America, 2017), pp. W2A–29.
21. C.-C. Chen, "Fast boundary detection: A generalization and a new algorithm," *IEEE Trans. Comput.* **C-26**(10), 988–998 (1977).
22. R. C. Gonzales and R. E. Woods, *Digital Image Processing* (Prentice hall, New Jersey, 2002).
23. J. W. Goodman, *Introduction to Fourier Optics* (W. H. Freeman, 2017), 4th ed.
24. Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microsc. Res. Tech.* **65**(3), 139–149 (2004).
25. X. Li, G. Liu, and J. Ni, "Autofocusing of ISAR images based on entropy minimization," *IEEE Trans. Aerosp. Electron. Syst.* **35**(4), 1240–1252 (1999).
26. Z. Ren, N. Chen, A. Chan, and E. Y. Lam, "Autofocusing of optical scanning holography based on entropy minimization," in *Digital Holography and Three-Dimensional Imaging*, (Optical Society of America, 2015), pp. DT4A–4.
27. M. Subbarao and J. K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Machine Intell.* **20**(8), 864–870 (1998).
28. M. Subbarao, "Focusing techniques," *Opt. Eng.* **32**(11), 2824 (1993).
29. P. Langehanenberg, B. Kemper, D. Dirksen, and G. Von Bally, "Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging," *Appl. Opt.* **47**(19), D176–D182 (2008).
30. F. C. Groen, I. T. Young, and G. Ligthart, "A comparison of different focus functions for use in autofocus algorithms," *Cytometry* **6**(2), 81–91 (1985).
31. I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proc. IEEE* **67**(5), 753–763 (1979).
32. J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vision* **31**(2/3), 111–127 (1999).
33. S. Lefkimiatis, A. Roussos, P. Maragos, and M. Unser, "Structure tensor total variation," *SIAM J. Imaging Sci.* **8**(2), 1090–1122 (2015).
34. J. Angulo, "Structure tensor image filtering using riemannian L1 and L $\infty$  center-of-mass," *Image Anal. Stereol.* **33**(2), 95–105 (2014).
35. W. Zhang, J. Fehrenbach, A. Desmaison, V. Lobjois, B. Ducommun, and P. Weiss, "Structure tensor based analysis of cells and nuclei organization in tissues," *IEEE Trans. Med. Imaging* **35**(1), 294–306 (2016).
36. S. Lefkimiatis, J. P. Ward, and M. Unser, "Hessian Schatten-norm regularization for linear inverse problems," *IEEE Trans. on Image Process.* **22**(5), 1873–1888 (2013).
37. G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3 (JHU press, 2012).
38. M. A. Bray, A. N. Fraser, T. P. Hasaka, and A. E. Carpenter, "Workflow and metrics for image quality control in large-scale high-content screens," *J. Biomol. Screen* **17**(2), 266–274 (2012).



39. V. Ljosa, K. L. Sokolnicki, and A. E. Carpenter, "Annotated high-throughput microscopy image sets for validation," *Nat. Methods* **9**(7), 637 (2012).
40. R. Liu, Z. Li, and J. Jia, "Image partial blur detection and classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2008), pp. 1–8.
41. J. Shi, L. Xu, and J. Jia, "Just noticeable defocus blur detection and estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2015), pp. 657–665.
42. F. Meng, Y. Ding, D. Wang, and J. Xiu, "Analysis of image motion on autofocus precision for aerial cameras," *Opt. Eng.* **54**(8), 083102 (2015).
43. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *IEEE International Conference on Computer Vision*, vol. 2 (IEEE, 2001), pp. 416–423.
44. S. Lefkimmiatis, A. Roussos, M. Unser, and P. Maragos, "Convex generalizations of total variation based on the structure tensor with applications to inverse problems," in *International Conference on Scale Space and Variational Methods in Computer Vision*, (Springer, 2013), pp. 48–60.