

# Single-frame rapid autofocusing for brightfield and fluorescence whole slide imaging

JUN LIAO,<sup>1</sup> LIHENG BIAN,<sup>2</sup> ZICHAO BIAN,<sup>1</sup> ZIBANG ZHANG,<sup>1,3</sup> CHARMI PATEL,<sup>1</sup> KAZUNORI HOSHINO,<sup>1</sup> YONINA C. ELДАР,<sup>4</sup> AND GUOAN ZHENG<sup>1,\*</sup>

<sup>1</sup>Biomedical Engineering, University of Connecticut, Storrs, CT 06269, USA

<sup>2</sup>Department of Automation, Tsinghua University, Beijing 100084, China

<sup>3</sup>Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China

<sup>4</sup>Electrical Engineering, Israel Institute of Technology, Haifa 32000, Israel

\*[guoan.zheng@uconn.edu](mailto:guoan.zheng@uconn.edu)

**Abstract:** A critical consideration for whole slide imaging (WSI) platform is to perform accurate autofocusing at high speed. Typical WSI systems acquire a z-stack of sample images and determine the best focal position by maximizing a figure of merit. This strategy, however, has suffered from several limitations, including low speed due to multiple image acquisitions, relatively low accuracy of focal plane estimation, short axial range for autofocusing, and difficulties in handling transparent samples. By exploring the autocorrelation property of the tissue sections, we report a novel single-frame autofocusing scheme to address the above challenges. In this approach, we place a two-pinhole-modulated camera at the epillumination arm. The captured image contains two copies of the sample separated by a certain distance. By identifying this distance, we can recover the defocus distance of the sample over a long z-range without z-scanning. To handle transparent samples, we set an offset distance to the autofocusing camera for generating out-of-focus contrast in the captured image. The single-frame nature of our scheme allows autofocusing even when the stage is in continuous motion. We demonstrate the use of our autofocusing scheme for fluorescence WSI and quantify the focusing performance on 1550 different tissue tiles. The average autofocusing error is  $\sim 0.11$  depth-of-field, 3 folds better than that of conventional methods. We report an autofocusing speed of 0.037 s per tile, which is much faster than that of conventional methods. The autofocusing range is  $\sim 80$   $\mu\text{m}$ , 8 folds longer than that of conventional methods. The reported scheme is able to solve the autofocusing challenges in WSI systems and may find applications in high-throughput brightfield/fluorescence WSI.

© 2016 Optical Society of America

**OCIS codes:** (180.0180) Microscopy; (170.0110) Imaging systems; (100.3010) Image reconstruction techniques.

## References and links

1. L. Pantanowitz, J. H. Sinar, W. H. Henricks, L. A. Fatheree, A. B. Carter, L. Contis, B. A. Beckwith, A. J. Evans, A. Lal, and A. V. Parwani; College of American Pathologists Pathology and Laboratory Quality Center, "Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center," *Arch. Pathol. Lab. Med.* **137**(12), 1710–1722 (2013).
2. J. R. Gilbertson, J. Ho, L. Anthony, D. M. Jukic, Y. Yagi, and A. V. Parwani, "Primary histologic diagnosis using automated whole slide imaging: a validation study," *BMC Clin. Pathol.* **6**(1), 4 (2006).
3. M. C. Montalto, R. R. McKay, and R. J. Filkins, "Autofocus methods of whole slide imaging systems and the introduction of a second-generation independent dual sensor scanning method," *J. Pathol. Inform.* **2**(1), 44 (2011).
4. S. Yazdanfar, K. B. Kenny, K. Tasimi, A. D. Corwin, E. L. Dixon, and R. J. Filkins, "Simple and robust image-based autofocusing for digital microscopy," *Opt. Express* **16**(12), 8670–8677 (2008).
5. L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, Jr., "Comparison of autofocus methods for automated microscopy," *Cytometry* **12**(3), 195–206 (1991).
6. R. R. McKay, V. A. Baxi, and M. C. Montalto, "The accuracy of dynamic predictive autofocusing for whole slide imaging," *J. Pathol. Inform.* **2**(1), 38 (2011).
7. K. Guo, J. Liao, Z. Bian, X. Heng, and G. Zheng, "InstantScope: a low-cost whole slide imaging system with instant focal plane detection," *Biomed. Opt. Express* **6**(9), 3210–3216 (2015).
8. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI*, 1981), 674–679.

9. Y. Fan, Y. Gal, and A. P. Bradley, "An algorithm for microscopic specimen delineation and focus candidate selection," *Micron* **66**, 51–62 (2014).

## 1. Introduction

Whole slide imaging (WSI) systems convert the conventional microscope slides into digital images that can be analyzed with computers and shared through the internet. It has become an important tool in biomedical research and clinical diagnosis [1]. In WSI imaging systems, autofocusing is the most challenging issue to overcome and has been cited as the culprit for poor image quality in histologic diagnosis [2]. This is not because autofocusing is difficult to do, but rather because of the need to perform accurate autofocusing at high speed [3]. There are two types of autofocusing methods: laser-reflection-based method and image-contrast-based method. Laser-reflection-based method cannot handle tissue sections with topography variations above the glass slide [3]. Conventional WSI systems use the image-contrast-based method to perform autofocusing [3–5]. This approach typically acquires multiple images by moving the sample (or the objective) along the axial direction and then selects the optimal focal plane by maximizing a figure of merit on the acquired images. Typical figures of merit include image contrast, resolution, entropy, and frequency content. The image-contrast-based method requires no reference surface and is able to track sample topography variations above the glass slide, making it a good solution for imaging tissue sections.

Despite its successful deployment in conventional WSI systems, the image-contrast-based approach suffers from several limitations: 1) it has a limited autofocusing speed due to the acquisition of multiple images per tile. Assuming a rate of 20 frames per second, surveying focus at 5 different focal positions per tile requires 0.25 seconds. This will be further limited by the motion of the stage in the  $z$  direction. Traditional tiling systems create a focus map by surveying every  $n$  tiles on the tissue. The assumption with skipping tiles is that a neighboring region has a similar focus position as its neighbors. More focus points increase the accuracy of the focus map while decreasing the speed. 2) It has a relatively low accuracy of focal plane estimation. It has been shown that the focusing error using a 3-point Brenner gradient method is about  $\sim 0.34$  depth of field (DOF) in a dynamic predictive mode [6]. 3) It has a relatively short axial range for autofocusing (typically  $< 10 \mu\text{m}$ ). If the sample is out of focus by a large amount, then it is difficult for image-contrast-based methods to recover the focal position. 4) Evident by its name, image-contrast-based technique relies on the image contrast of the captured data. Thus, it is difficult to handle unstained, transparent, or low-contrast samples. It is unclear whether image-contrast-based methods can be implemented for fluorescence microscopy, where samples are typically transparent under brightfield illumination. One can use a fluorescence channel for obtaining image contrast; however, capturing multiple low-light fluorescence images for autofocusing may be time-consuming and introduces photobleaching damages to the samples.

In this work, we report a novel, robust, and rapid autofocusing approach based on single image acquisition. Our setup integrates the dual-camera configuration [3] and the pinhole-modulation idea [7] to address the challenges discussed above. Different from the original pinhole-modulation idea of using two images, the reported scheme only need to capture one image for autofocusing. The eyepiece ports are also released for clinicians' use. More importantly, the original pinhole-modulation scheme cannot be used for fluorescence imaging. The reported scheme, on the other hand, is able to handle transparent samples and be used for both brightfield and fluorescence WSI. The single-frame nature of the reported scheme also allows autofocusing even the stage is in continuous motion. The average autofocusing error of the reported scheme is  $\sim 0.11$  depth-of-field,  $\sim 3$  folds better than that of conventional image-contrast-based methods. The time to determine the best focus position is 0.037 seconds, much faster than that of conventional methods. The autofocusing range is  $\sim 80 \mu\text{m}$ , 8 folds longer than that of conventional methods. The reported scheme may find applications in high-throughput WSI and DNA-sequencing.

## 2. Single-frame rapid autofocusing scheme

The reported single-frame autofocusing technique is inspired by the dual-camera configuration, where the high-speed camera is used for autofocusing and the main camera is used for capturing high-resolution images [3]. As shown in Fig. 1(a), we placed the autofocusing camera module at the epi-illumination arm. This module consists of a filter cube, two 50-mm CCTV lenses, a two-pinhole aperture at the pupil plane, and a cost-effective image sensor (Sony IMX265). In this setup, we used a surface-mount LED (LOHAS 50W LED) for sample illumination, which was placed at the back focal plane of the condenser lens. Figure 1(a3) shows the entire WSI platform, where we used three stepping motors to control the motion of the microscope stage in the x, y, and z directions [7]. In the reported autofocusing scheme, the light from the sample is divided into two paths by the beam splitter: one goes to the high-resolution main camera at the top and the other goes to the autofocusing camera. By placing the two-pinhole aperture at the pupil plane, the captured image from the autofocusing camera contains two copies of the sample and the translational shift of these two copies is proportional to the defocus distance (Fig. 1(b1)-1(b3)). Figure 1(b4) shows the relationship between the translational shift of the two copies and the defocus distance (the three color data points in Fig. 1(b4) correspond to the cases of Fig. 1(b1)-1(b3)). Once we identify the translation shift between the two copies, we can recover the defocus distance based on the curve in Fig. 1(b4). In our implementation, we used 2 by 2 binning for the autofocusing camera and the captured image contains 1024 by 768 pixels. We used the central 768 by 768 region for processing. We note that we have set up an offset for the autofocusing camera in our platform; in other words, when the sample is in-focus, there is a translational shift of the two copies (Fig. 1(b2)). This offset is able to generate out-of-focus contrast for the transparent sample, as evident in Fig. 1(b1)-1(b3) and the inset of Fig. 1(b4). We will further discuss this point below.

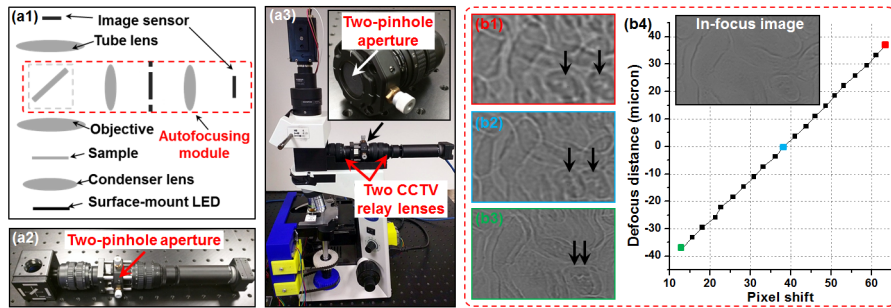


Fig. 1. The single-frame autofocusing scheme. (a) The microscope setup, where the autofocusing module is attached at the epi-illumination arm. (b) The working principle of the single-frame autofocusing scheme. The captured image from the autofocusing camera contains two copies of the object and we can recover the defocus distance based on the translation shift between the two copies.

The first question is how to recover the translational shift from the single captured image. This problem is different from the shift retrieval problem in stereo vision, where phase correlation can be calculated from two images [8]. In our case, we have one measurement  $z[x] = s[x] + s[x - x_0]$ , where  $s[x]$  and  $s[x - x_0]$  represent two copies of the sample in Fig. 1(b). The goal is to recover the shift  $x_0$  from  $z[x]$  ( $s[x]$  is unknown).

We first rewrite  $z[x]$  as follows:  $z[x] = s[x] + s[x - x_0] = s[x] * h[x]$ , where  $h[x] = \delta[x] + \delta[x - x_0]$  and  $*$  stands for convolution. We propose to recover  $x_0$  from the autocorrelation of the captured image  $z[x]$ . Specifically, the autocorrelation of  $z[x]$  can be expressed as

$$R(z[x]) = R(s[x]) * R(h[x]) = R(s[x]) * (2\delta[x] + \delta[x - x_0] + \delta[x + x_0]), \quad (1)$$

where ‘ $R()$ ’ stands for the autocorrelation operation. The term ‘ $2\delta[x] + \delta[x - x_0] + \delta[x + x_0]$ ’ in Eq. (1) suggests that if  $R[s[x]]$  is narrow enough, then there will be three peaks in the autocorrelation function  $R(z[x])$ , one at the center, one at the  $x_0$  position, and one at the  $-x_0$  position. Therefore, in this case, we can recover  $x_0$  by identifying the locations of the two first-order peaks of  $R(z[x])$ .

By definition, the autocorrelation function  $R(z[x])$  can be computed by a convolution operation:  $R(z[x]) = z[x] * z[-x]$ . In practice, the Wiener-Khinchin theorem allows us to compute  $R(z[x])$  with two fast Fourier transforms (FFTs): first compute the Fourier power spectrum of the captured image  $z[x]$  and then perform an inverse FFT on the power spectrum. Figure 2 summarizes the procedures: we first compute the Fourier power spectrum in Fig. 2(a2) and then perform an inverse FFT to get the autocorrelation function  $R(z[x])$  in Fig. 2(a3). The distance  $x_0$  can be recovered from the distance between the two first-order peaks in Fig. 2(a4).

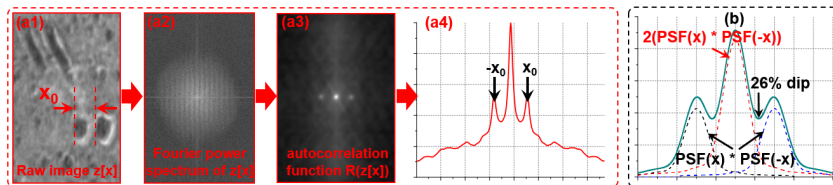


Fig. 2. The procedures for recovering the translation shift from a single captured image  $z[x]$ . (a1) The captured image  $z[x]$  from the autofocusing camera. (a2) The Fourier power spectrum of the captured image (we took the log scale to better visualize the fringe pattern). (a3) The autocorrelation function  $R(z[x])$ , which can be computed by taking the inverse Fourier transform of (a2). (a4) The line trace of (a3) and the locations of the peaks. (b) The condition for resolving the first-order peaks.

Although the procedures in Fig. 2 works well in many cases, we cannot guarantee that it will always recover  $x_0$ . To gain more intuition into the method, consider two extreme cases for  $s[x]$ : 1)  $s[x]$  is a constant, and 2)  $s[x]$  is an i.i.d. random function. For case 1, the correlation of a constant is still a constant. Therefore, we will get 3 constants overlapped with each other from Eq. (1) and we cannot recover the distance  $x_0$ . For case 2, the correlation function will be a  $\delta$  function so that Eq. (1) leads to 3  $\delta$  functions. We can, therefore, recover  $x_0$  from the locations of the  $\delta$  functions. In practice, a good model for  $s[x]$  is a broadband object  $o[x]$  (with narrow correlation function) convolved with the incoherent point spread function (PSF) of the imaging system. Therefore, the power spectrum of  $s[x]$  can be approximated by a constant times the magnitude squared of OTF, where ‘OTF’ stands for the optical transfer function (i.e., the Fourier transform of the PSF). Equation (1) then leads to three copies of the correlation function of the PSF in Fig. 2(b). We can then define the following condition for resolving the locations of the first-order peaks: the dip adjacent to the first-order peak is at least 26% lower than the peak value. A similar condition is used in the Rayleigh criterion for defining the resolution of two closely-packed peaks. Under the condition in Fig. 2(b), we can get the following important requirement on  $x_0$ :

$$x_0 \cdot f_{\text{cutoff}} > 1.56, \quad (2)$$

where  $f_{\text{cutoff}}$  stands for the cutoff frequency of the incoherent OTF and is equal to  $2NA/\lambda$  for an aberration-free system. Equation (2) implies that, if the distance between the two copies is small, then it will be difficult to recover  $x_0$ . This observation justifies the positional offset of the autofocusing camera in our platform. We set this offset for two purposes: 1) to generate out-of-focus contrast for the captured image, and 2) to satisfy Eq. (2). We also note that the auto-phase correlation index can be used in the acquisition process to select focus candidates [9].

### 3. Autofocusing performance and fluorescence WSI

In Fig. 2(a4), we need to identify the locations of the two first order peaks to recover  $x_0$ . A simple solution is to locate the local maximum point, as shown by the black arrow in Fig. 3(a1). This solution leads to the step-wise relationship between the recovered  $x_0$  and the defocus distance, as shown by the black curve of Fig. 3(a2). This behavior is due to the limited precision of the recovered  $x_0$ . To achieve sub-pixel precision, we can perform curve fitting to better identify the locations of the first-order peaks. For the red curve in Fig. 3(a1), we used a 5-point smoothing spline fitting to estimate the locations of the first-order peaks. The resulting relationship between  $x_0$  and the defocus distance is shown in the red curve of Fig. 3(a2), where we can see a linear relationship between the two.

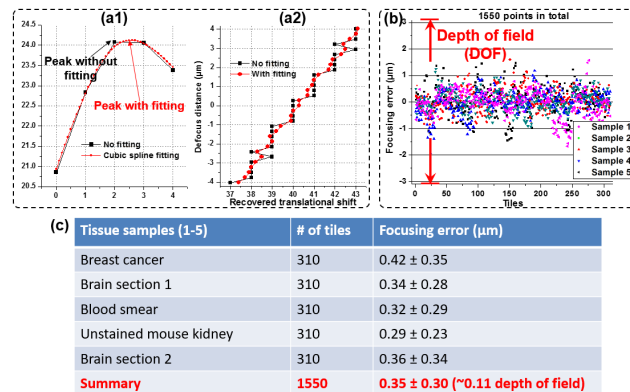


Fig. 3. The autofocusing performance of our scheme. (a) Achieving a sub-pixel accuracy of the translational shift estimation. (b) The focusing error on 5 samples and 1550 different tiles. (c) Summary of the autofocusing performance. We used a 10-point Brenner gradient method to determine the ground truth position. The average focusing error is  $\sim 0.11$  DOF,  $\sim 3$  folds better than the conventional image-contrast-based method.

To quantify the performance of the reported scheme, we tested the platform on 5 different tissue sections and 1550 different tiles. The stage is fixed during the autofocusing operation and the camera offset is chosen for achieving a  $\sim 80 \mu\text{m}$  autofocusing range. Figures 3(b) and 3(c) summarize the results. In particular, the time to determine the best focus position (from image acquisition to the output of the defocus position) is  $\sim 0.037$  s, much faster than that of conventional image-contrast-based methods; 45% of the 0.037-s duration is consumed by the two fast Fourier transform (FFT) operations in Fig. 2. Therefore, the speed can be further improved using parallel computing techniques or an FPGA. Figure 3(b) shows the focusing error for the 1550 tissue tiles using a 20X 0.4 NA objective lens, with a depth-of-field (DOF) of  $\pm 3.125 \mu\text{m}$ . The average focusing error is  $\sim 350$  nm, which is  $\sim 0.11$  DOF. In contrast, the average focusing error of the 3-point Brenner gradient method is  $\sim 0.34$  DOF in a dynamic predictive mode and  $\sim 0.2$  DOF in a static mode [6]. Our approach is  $\sim 3$  folds better than that of the dynamic predictive mode and  $\sim 2$  folds better than that of the static mode. In addition, both stained and transparent samples have similar performance in our scheme.

For fluorescence WSI, two strategies can be used for autofocusing. The first one is to acquire a z-stack of fluorescence images and determine the best focus position using the Brenner gradient method. The acquisition of multiple fluorescence images, however, may be extremely time-consuming and introduce photobleaching to the sample. The second strategy is to use the brightfield channel for autofocusing and then acquire the fluorescence image, as suggest by Ref [4]. This strategy, however, may be problematic as many fluorescence samples are transparent under brightfield illumination. It only works for samples with both brightfield and fluorescence staining. To the best of our knowledge, the reported scheme is the first effective approach for both brightfield and fluorescence WSI. It uses the unwanted

brightfield channel for autofocusing, and thus, no fluorescence photon is lost in the acquisition process. It can handle transparent samples by introducing an offset to the autofocusing camera. Figure 4 shows the whole slide fluorescence images captured by using the reported platform.

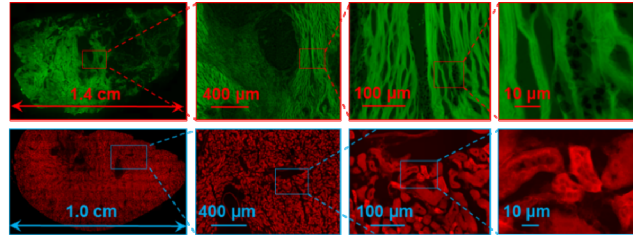


Fig. 4. The fluorescence images of a breast cancer (top) and an unstained mouse kidney section (bottom). The full images can be found from <http://gigapan.com/profiles/SmartImagingLab>.

#### 4. Summary

We have reported a novel autofocusing scheme for brightfield and fluorescence whole slide imaging. In our approach, we place a two-pinhole-modulated camera at the epi-illumination arm. The captured image contains two copies of the sample separated by a certain distance. By identifying this distance, we can recover the defocus distance of the sample over a long z-range and without z-scanning. We have also discussed conditions for recovering the distance between the two copies. In particular, we introduce a positional offset to the autofocusing camera to satisfy the autofocusing condition in Eq. (2) and to generate out-of-focus image contrast.

There are several important advantages to the suggested scheme: 1) it only needs one image for autofocusing, and thus, it shortens the time for producing a focus map in WSI platforms. More importantly, the single-frame nature of the reported scheme allows autofocusing even when the stage is in continuous motion (with pulsed illumination). The use of single image for autofocusing is a clear advantage over the dual-camera technique reported in Ref [3], where rapid z-scanning is needed for each tile. The speed for autofocusing speed is 0.037 s per tile, which is, to the best of our knowledge, a record-high speed. 2) The autofocusing performance is  $\sim 3$  folds better than that of image-contrast-based methods. 3) The autofocusing range is at least 80  $\mu\text{m}$  in the reported prototype platform and it is  $\sim 8$  folds better than that of conventional approaches. 4) The reported scheme is able to handle transparent or unstained samples, which is a clear advantage over other existing methods. 5) Our approach requires only a cost-effective microscope add-on kit as shown in Fig. 1(b2). The dissemination of the proposed scheme for WSI brightfield and fluorescence imaging under a limited budget will enable new types of experimental designs in biological and clinical labs, e.g., digital pathology, cytology analysis, genetic studies on multicellular organisms, drug profiling, DNA sequencing, and more.

One future direction is to investigate the optimal mask placed at the Fourier plane. The two-pinhole mask may not be optimal for recovering the defocus distance. Effort along this direction is on-going. Another direction is to implement pulsed illumination, which allows autofocusing while the stage is in continuous motion. Performing accurate autofocusing at high speed is the Achilles' heel of WSI. The reported scheme may provide a transformative solution for brightfield/ fluorescence WSI, in particular, for handling transparent and low-contrast samples.

#### Acknowledgment

This work was supported by NSF DBI 1555986. L. Bian acknowledges the support of the National Natural Science Foundation of China under grant 61327902. Z. Zhang acknowledges the support of the China Scholarship Council.