Computational imaging modalities for multi-focal whole-slide imaging systems

Van Der Graaff, Leon; Van Leenders, Geert J.L.H.; Boyaval, Fanny; Stallinga, Sjoerd

DOI
10.1364/AO.394290

Publication date
2020

Document Version
Final published version

Published in
Applied Optics

Research Article Vol. 59, No. 20 / 10 July 2020 / Applied Optics 5967

applied optics

Computational imaging modalities for multi-focal whole-slide imaging systems

Leon van der Graaff,[1,4] Geert J. L. H. van Leenders,[2] Fanny Boyaval,[3] and Sjoerd Stallinga[1,5]

[1]Department of Imaging Physics, Delft University of Technology, The Netherlands
[2]Department of Pathology, Erasmus University Medical Center, The Netherlands
[3]Department of Pathology, Leiden University Medical Center, The Netherlands
[4]e-mail: l.vandergraaff@tudelft.nl
[5]e-mail: s.stallinga@tudelft.nl

Received 8 April 2020; revised 8 June 2020; accepted 8 June 2020; posted 9 June 2020 (Doc. ID 394290); published 6 July 2020

Whole-slide imaging systems can generate full-color image data of tissue slides efficiently, which are needed for digital pathology applications. This paper focuses on a scanner architecture that is based on a multi-line image sensor that is tilted with respect to the optical axis, such that every line of the sensor scans the tissue slide at a different focus level. This scanner platform is designed for imaging with continuous autofocus and inherent color registration at a throughput of the order of 400 MPx/s. Here, single-scan multi-focal whole-slide imaging, enabled by this platform, is explored. In particular, two computational imaging modalities based on multi-focal image data are studied. First, 3D imaging of thick absorption stained slides (∼60 μm) is demonstrated in combination with deconvolution to ameliorate the inherently weak contrast in thick-tissue imaging. Second, quantitative phase tomography is demonstrated on unstained tissue slides and on fluorescently stained slides, revealing morphological features com-plementary to features made visible with conventional absorption or fluorescence stains. For both computational approaches simplified algorithms are proposed, targeted for straightforward parallel processing implementation at ∼GPx/s throughputs. © 2020 Optical Society of America

https://doi.org/10.1364/AO.394290

1. INTRODUCTION

A higher resolution and a larger field of view (FOV) have always been an important concern in the design of optical systems. The use of an objective lens with a higher numerical aperture (NA) provides a higher resolution but at the same time decreases the FOV of the optical system. Whole-slide imaging (WSI) systems provide a solution to this fundamental trade-off between resolution and FOV by scanning the tissue slide and acquiring a digital image with high resolution (∼0.25 μm) with a truly unlimited FOV. WSI systems are the primary enabling technology for digital pathology, with applications in primary diagnosis [1–4], education [5–7], and research [8,9].

Other methods are proposed to overcome the trade-off between resolution and FOV. Multi-spot scanning with array illuminators—here a large grid of high-NA spots is used for illumination, but a low NA with large FOV imaging path is used—have been considered [10–12]. In Fourier ptychography [13,14], a series of low-resolution images of a large FOV acquired with different illumination beam angles are combined to produce a single high-resolution image of the entire FOV. Although these methods do overcome the resolution–FOV trade-off, they do not optimally use the space-bandwidth-time product of the optical system [15] (basically corresponding to the throughput in Mpx/s) because a fraction of the data stream is lost to overhead or carries redundant information. In addition, non-scanning approaches are still limited in FOV by the low-magnification objective lenses used.

Currently, the majority of WSI systems provides high-quality two-dimensional (2D) images of tissue slides, mostly for brightfield microscopy with white-light illumination [16–18]. There are two main methods of scanning. The first is the "step -and-stitch" method, in which the slide is moved step-wise and imaged using an area scanner. This is the most trivial extension to the traditional FOV-limited microscope but requires a potentially complex process of image-stitching in post-processing. The second is the "push broom" or line scanning method, which combines simplicity and speed by scanning the slide with a line sensor at a constant velocity. The limited illumination etendue can be compensated by sensors with a time delay integration (TDI) capability [19]. Current scanner systems have a throughput of several 100 Mpx/s, sufficient for imaging ∼cm^2 tissue areas at submicrometer sampling density within ∼1 min.

A major challenge in slide scanning is the need for tracking the topographic variations of the tissue layer which usually

exceed the depth-of-focus of the scanner. To this end, a focus map of the whole slide can be made prior to scanning, where only a limited number of locations is used to save time [20]. An improvement is an autofocus system in which focus information is provided continuously, without mechanically changing focus. This information is then used for a closed-loop feedback system for maintaining optimum focus [21,22]. Such an autofocus system can be realized using a second optical branch in the scanner for generating the focus information. Drawbacks of this approach are the increase in overall optical complexity and the required synchronization between the autofocus and imaging branches of the scanner.

Recently, a new WSI system has been described in the patent literature by Philips [23–25] to overcome these drawbacks. The core of the approach lies in a new architecture for a multi-line image sensor. The sensor contains separate "sensorlets," groups of adjacent pixel rows, which can be read-out independently. The sensor is tilted with respect to the optical axis resulting in a tilted object plane [26]. The readout of pixel information is done via two separate, simultaneously obtained, data streams, one obtained from a single sensorlet for providing the primary imaging information, the other obtained from multiple sensorlets for providing autofocus information. A full-color whole-slide image with inherent color registration is acquired with a color-sequential illumination scheme based on fast switchable, high-power LEDs [27]. In summary, the advantages of the proposed architecture are the elimination of the need for multiple cameras for imaging and autofocusing, the reduction of component costs, and the simplification of manufacturing and maintenance.

An inherent capacity of the WSI platform is the possibility to make a multi-focal scan by readout of multiple sensorlets simultaneously [23–25]. This reduces scan time and has intrinsic axial registration of the scanned layers compared to acquiring a multi-layer image with multiple sequential 2D scans. The goal of this paper is to explore new imaging modalities of the scanner platform, based on this opportunity. Making 3D images with a single scan specifically leads to opportunities in imaging the 3D morphology of tissues and cells over large FOVs and in computational phase contrast imaging of unstained tissues.

Visualizing the 3D tissue morphology adds value in some cases in the field of histology for making a final diagnosis [16,28]. Visualizing cells in 3D is generally always needed for diagnoses in the field of cytology [29]. The 3D imaging functionality of current WSI systems is mostly realized through sequential scans at different focus levels [30–32]. This approach leads to a considerable increase in scan time, and is sensitive to errors in registration of focal layers, which hamper its large-scale application. These disadvantages can be overcome by the proposed single-scan 3D imaging mode.

Quantitative phase imaging forms a non-invasive and label-free imaging platform in cell biology and pathology [33]. Algorithms are available for phase retrieval from a through-focus image stack to obtain a 2D phase contrast image of a thin layer, usually based on solving the transport of intensity equation (TIE) [34–37], as well as for a full 3D tomographic reconstruction of a thicker specimen [38–40]. Application of such computational phase contrast modalities based on multi-focal image stacks can broaden the application of WSI systems to unstained samples.

An important requirement for both modalities is the need for simple, scalable image processing methods, ultimately enabling real-time image data processing. Key in achieving high-throughput image processing is parallelization of the algorithm, such that it can be computed efficiently on a graphics processing unit (GPU) or on dedicated hardware, e.g., a field-programmable gate array (FPGA). In this work the focus will therefore be on image processing algorithms that are compatible with large-scale parallelization in a straightforward way. In particular, a deconvolution algorithm for improved sectioning in multi-focal volumetric imaging of thick slides and an algorithm for quantitative phase tomography are presented. In both a "z-only" approach is proposed, in which the final deconvolved or phase image data is calculated independently for each lateral position (x, y), which leads to low memory, efficient, and scalable calculations.

This paper is structured as follows. First, the 3D scanner architecture is described including the novel image sensor. Next, multi-focal volumetric images of 60 μm thick tissue sections are shown. Then, the results of quantitative phase tomography of unstained tissue sections are presented. The paper is concluded with an evaluation of the findings and possible next steps for the highly modular WSI platform.

2. SCANNER SETUP AND IMAGE PROCESSING

A. 3D Scanner Architecture

A schematic outline of the 3D WSI platform is shown in Fig. 1. At the heart of the system lies a CMOS image sensor, of which the pixel geometry is outlined in Fig. 1(a). The sensor consists of 128 sensorlets, groups of four adjacent pixel rows with a width of 4096 pixels. The sensorlets have a 13-row spacing. The pixel pitch is $\Delta_l = 6.42$ μm in the scan direction and $\Delta_p = 5.54$ μm in the orthogonal, or "field" direction. The sensorlets have a 13-row spacing such that the sensorlet pitch $\Delta_s = (13 + 4)\Delta_l = 115.56$ μm. The sensor is tilted over an angle of $\beta = 30$ deg, so that each sensorlet scans the sample at a different depth; see Fig. 1(e). This also results in a square projection of the pixels on the plane of the tissue slide. The slide is illuminated by a color-sequential LED-based light source, equipped with a Köhler condenser [27]; see Fig. 1(f). The light source has three color channels for RGB imaging, with a typical wavelength of 657 nm for the red channel, 557 nm for the green channel, and 465 nm for the blue channel.

Two FPGA modules or "engines" govern the readout of the sensor and provide the capability of on-chip processing of the image data. Each engine has four analog-to-digital converters (ADCs), which can be independently connected to a sensorlet of choice. The main purpose of having two engines is to use one for acquiring the image data, while the other engine is used to capture autofocus data, as is illustrated in Fig. 1(b). Optionally, the readout signal can be increased with a factor of 4 by application of TDI [19]. In this case, all four ADCs are connected to the four rows of a single sensorlet, and the FPGA applies digital TDI to create the primary, high-resolution, high-SNR, image information. Simultaneously, the other engine reads out a series of sensorlets sequentially, e.g., starting at sensorlet

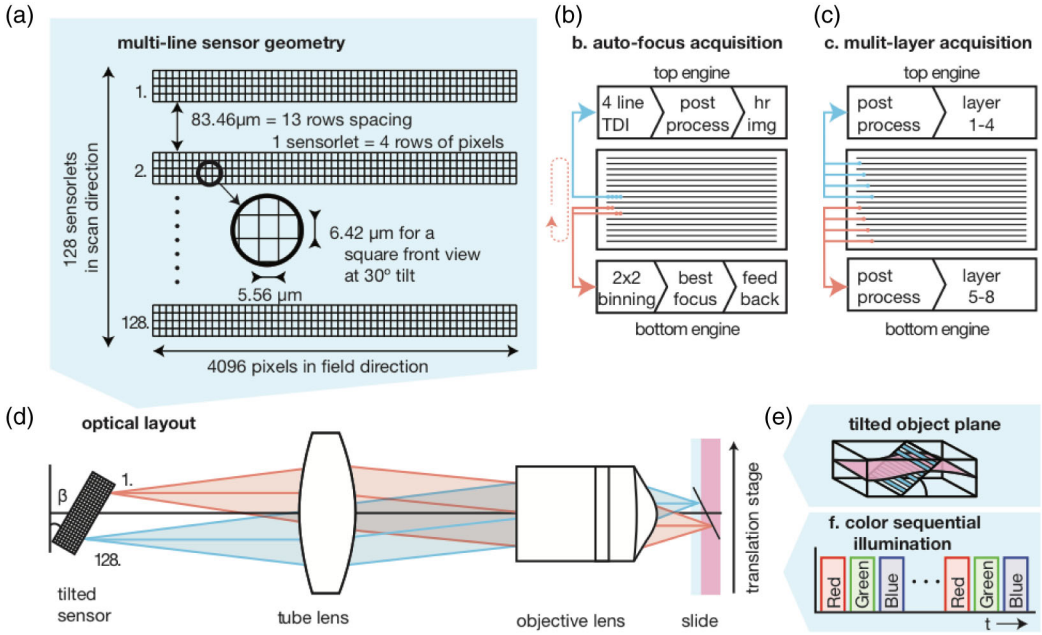

Fig. 1. Schematic layout of the WSI system. (a) The pixel geometry of the novel multi-line CMOS image sensor. The sensor is equipped with two "engines" that can independently read out the sensor. This can be used for (b) 2D scanning with continuous autofocus or (c) 3D scanning. The optical layout (d) shows the tilted sensor that gives rise to (e) a tilted object plane. (f) A color sequential illumination strategy is used for RGB imaging.

one, and increasing the sensorlet number with every readout line. Optionally, this is combined with "binning" for higher signal levels, where the FPGA combines the data of two rows of a sensorlet to create a two times down-sampled image with a four times higher signal level. The FPGA can calculate the optimal focus position from this data, and provides real-time feedback to adjust the position of the objective lens. In this study we use an alternative readout mode in which data is acquired from eight different sensorlets simultaneously, as is illustrated in Fig. 1(c). For example, by reading out a single row of the sensorlets 1, 17,...,127, an eight-layer multi-focal volume covering the full axial range can be obtained.

Images are acquired in a line-by-line or "push broom" scanning fashion [41,42]. The translation stages perform a linear translation with a velocity v_t while the sensor is triggered for data acquisition at equidistant positions with a step size Δ_t. We use $\Delta_t = 0.25$ μm and $v_t = 1$ mm/s resulting in a throughput of $4096 \cdot 3 \cdot 8 \cdot v_t/\Delta_t \approx 393$ MPx/s.

The objective lens and tube lens form a telecentric imaging system; see Fig. 1(d). We use a Nikon 20 × NA.75 Plan Apochromat VC objective lens ($F_{\text{obj}} = 10$ mm) for high-resolution imaging. Alternatively, a Nikon 10 × NA.45 CFI Plan Apochromat λ objective lens ($F_{\text{obj}} = 20$ mm) is used for a 2× wider scan lane and larger axial range, at the expense of lateral resolution. The tube lens has an effective back focal length of $F_{\text{tube}} = 222.4 \pm 2.2$ mm to match the sensor pixel pitch of 5.54 μm with the target sampling density of $\Delta = 0.25$ μm.

Scanning experiments were carried out on a prototype WSI system realized by Philips for research purposes.

B. Finite Conjugate Imaging and Spherical Aberration

A range of axial positions in object (sample) space is imaged onto the tilted detector by the imaging light path, comprising the objective lens and the tube lens. This imaging light path is optimized for imaging at a single depth inside the sample, typically directly after the coverslip. It follows that in this case we will suffer from spherical aberration. The sensitivity to spherical aberration can be analyzed along the lines of [43]. In that analysis, it is assumed that the objective lens and tube lens form an aplanatic telecentric imaging system.

We consider three degrees of freedom (see Fig. 2), namely z_1, the axial object position; z_2, the increase of the free working distance of the objective with respect to the nominal working distance; and z_3, the axial image position. These three distances are relative to the nominal aberration-free case. The total aberration function $W(\rho)$ is then given by

$$W(\rho) = z_1\sqrt{n^2 - \rho^2\text{NA}^2} + z_2\sqrt{1 - \rho^2\text{NA}^2} - z_3\sqrt{1 - \frac{\rho^2\text{NA}^2}{M^2}}, \tag{1}$$

where ρ is the scaled radial pupil coordinate ($0 \le \rho \le 1$) such that the pupil is scaled to the unit circle, NA denotes the objective numerical aperture, $M = F_{\text{tube}}/F_{\text{ob}}$ is the lateral magnification, equal to the ratio of the tube focal length and the objective focal length, and n is the medium refractive index. This may be written in a compact form as

$$W(\rho) = \sum_{j=1}^{3} z_j f_j(\rho), \tag{2}$$

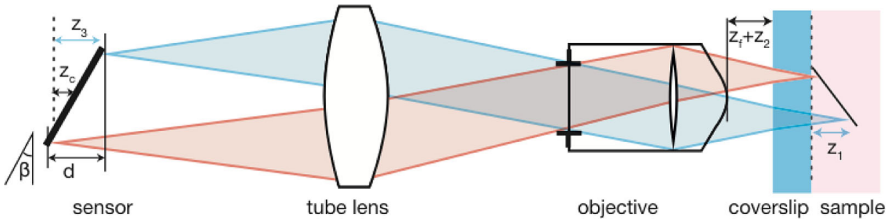

Fig. 2. Illustration of the effect of the shifted axial position on focus and conjugate for two lines on the sensor (in blue and red). The light path runs from right to left so as to be consistent with previous figures. The sample consists of a coverslip (cyan) and the tissue of interest (pink) which are assumed to have a matched refractive index n. The dashed lines indicate the nominal object and image planes, i.e., the focal plane of the tube lens and the surface below the coverslip. Three degrees of freedom are indicated: the axial object position z_1, the free working distance $z_f + z_2$ where z_f is the nominal free working distance, and the axial image position z_3. Tilting the sensor over an angle β makes it cover an axial range d in image space. Changing the axial position z_c of the sensor is used for minimizing the overall spherical aberration.

$$f_1(\rho) = \sqrt{n^2 - \rho^2 \mathrm{NA}^2}, \tag{3}$$

$$f_2(\rho) = \sqrt{1 - \rho^2 \mathrm{NA}^2}, \tag{4}$$

$$f_3(\rho) = \frac{\rho^2 \mathrm{NA}^2}{2M^2}, \tag{5}$$

where a Taylor approximation is used for the third aberration term based on $M \gg 1$ and neglecting the piston term. The total amount of aberration is given by the root mean square (rms) value of the aberration function W_{rms}, which is given by

$$W_{\mathrm{rms}}^2 = \sum_{j,l=1}^{3} g_{jl} z_j z_l, \tag{6}$$

with

$$g_{jl} = \langle f_j f_l \rangle - \langle f_j \rangle \langle f_l \rangle. \tag{7}$$

Here the angular brackets indicate averaging over the pupil (i.e., integration over the unit circle with radial coordinate ρ). These averages can be evaluated analytically and are given in [43].

The degrees of freedom in our system can be reduced using two conditions. First, the object should be in focus. This is defined by the axial object position z_1 for which the rms value of the aberration function is minimal, given the axial position of the sensor z_3 and the objective's working distance z_2. This implies that z_1 can be found by solving

$$\frac{\partial W_{\mathrm{rms}}}{\partial z_1} = 0, \tag{8}$$

which leads to

$$z_1 = -\frac{g_{12}}{g_{11}} z_2 - \frac{g_{13}}{g_{11}} z_3. \tag{9}$$

This expression directly gives the axial magnification as

$$M_{\mathrm{ax}} = -\frac{g_{11}}{g_{13}} = \chi \frac{M^2}{n}, \tag{10}$$

where χ is a non-paraxial correction factor depending only on NA and n, which satisfies $\chi \to 1$ in the limit $\mathrm{NA}/n \to 0$.

The second condition we impose is that the upper focal slice should be adjacent to the coverslip. The tilted image sensor spans a range of axial positions $z_c - d/2 \le z_3 \le z_c + d/2$ in image space, where d is the total axial range and z_c is the axial position of the center of the sensor. In object space, this corresponds to the axial range $0 \le z_1 \le d/M_{||}$. Now Eq. (9) implies that the working distance of the objective must be set such that

$$z_2 = -\frac{g_{13}}{g_{12}} \left(z_c - \frac{d}{2} \right). \tag{11}$$

Using the expressions we derived for the object axial position and the working distance [see Eqs. (9) and (11)], we can now write the rms value of the aberration function as a function of the axial image position, in which the axial position for the center of the sensor z_c remains as a free parameter. With some algebra it may be shown that

$$W_{\mathrm{rms}}^2 = \left(g_{33} - \frac{g_{13}^2}{g_{11}} \right) z_3^2 + 2 \left(\frac{g_{13}^2}{g_{11}} - \frac{g_{23} g_{13}}{g_{12}} \right) \left(z_c - \frac{d}{2} \right) z_3 + \left(\frac{g_{22} g_{13}^2}{g_{23}^2} - \frac{g_{13}^2}{g_{11}} \right) \left(z_c - \frac{d}{2} \right)^2. \tag{12}$$

Clearly, the aberration depends quadratically on the axial image position z_3, which is illustrated in Fig. 3. We can now choose z_c to minimize the overall aberration. This is achieved when the minimum of the parabola is at $z_3 = z_c$. This condition can be used to find an expression for z_c:

$$z_c = \frac{g_{13}^2/g_{11} - g_{23} g_{13}/g_{12}}{g_{33} - g_{23} g_{13}/g_{12}} \frac{d}{2}. \tag{13}$$

We will discuss the implications for the default $20 \times$ NA0.75 objective lens. The results for other optical configurations are given in Table 1. Using this objective, the lateral magnification is $M = 22.2$. The sensor has a pixel pitch of 5.56 μm, which gives a pixel pitch in object space equal to $p = 0.25$ μm. Assuming a medium refractive index $n = 1.5$, it is found that $\chi = 1.07$ and the axial magnification is $M_{\mathrm{ax}} = 352$. The image sensor size in the direction perpendicular to the line sensors is $127 \times 17 \times 6.42$ μm $= 13.9$ mm. With a tilt angle $\beta = 30$ deg the axial range in image space is $d = 6.9$ mm and the axial range

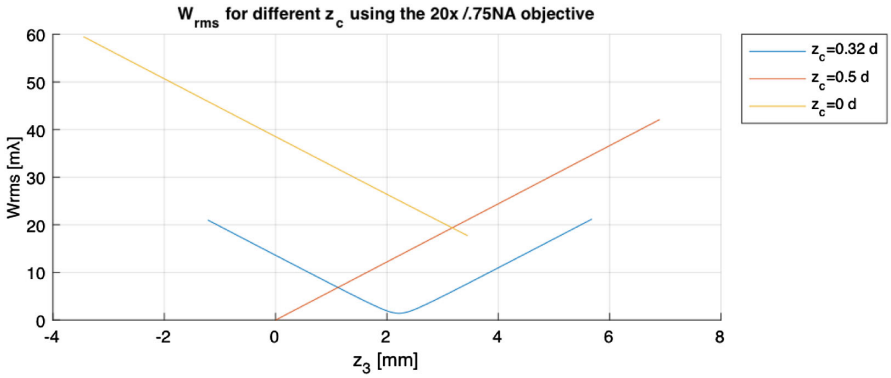

Fig. 3. Total aberration W_{rms} as a function of the axial position in image (detector) space for different values of z_c: the optimal value ($z_c = 0.32d$), with the top of the sensor at the nominal plane ($z_c = d/2$), and with the center of the sensor at the nominal plane ($z_c = 0$). The curves are calculated using NA = 0.75 and $\lambda = 500$ nm.

Table 1. Calculated Optical Properties for Optical Configurations, with Different Magnification and Axial Range of the 3D Scan[a]

Obj.	F_{obj} [mm]	M	M_{ax}	χ	Δ [μm]	Δ_{ax} [μm]	d/M_{ax} [μm]	W_{rms} [$m\lambda$]
20 × /.75	10	22.2	352	1.07	0.25	0.15	19.7	1.4–21
10 × /.45	20	11.1	84.1	1.02	0.50	0.65	82.4	0.2–10
5 × /.20	50	5.55	13.2	1.00	1.25	2.65	512	0.0–2.4

[a]Shown are the used objective lens, the NA, the focal length F_{obj} of the objective lens, the lateral magnification M, the axial magnification M_{ax}, the non-paraxial correction factor χ, the lateral pixel pitch in object space Δ, the target sampling density in object space Δ_{ax}, the resulting axial range in object space d/M_{ax}, and the root mean square spherical aberrations W_{rms} at the center and edge of the sensor, respectively. The values are based on $n = 1.5$, $\lambda = 500$ nm, and $\beta = 30$ deg.

in object space is $d/M_{\text{ax}} = 19.7$ μm. The optimal axial position of the center of the sensor is $z_c = 0.32d$, giving an axial range in image space of $-0.18d \le z_3 \le 0.82d$. The reason for the significant asymmetry is the use of finite conjugate compensation of the spherical aberration arising from focusing into the sample below the coverslip.

Figure 3 shows the rms spherical aberration as a function of the axial position in image (detector) space. With the optimum choice for z_c, the rms spherical aberration for the center of the sensor is only 1.4 mλ for a typical green emission wavelength of $\lambda = 500$ nm, composed mainly of higher-order spherical aberration, and 21 mλ at the edge of the sensor. The spherical aberration varies in an asymmetrical way from the bottom of the sensor to the top of the sensor, in case the axial position of the middle of the sensor is not chosen optimally. For example, by choosing $z_c = 0$, the center of the sensor is at the nominal image plane. Now the rms spherical aberration ranges up to 60 m λ for the worst-case edge of the sensor. When $z_c = d/2$ is chosen, the top row of the sensor is in the nominal image plane and has no aberration. However, the row at the other side of the sensor will then experience an rms aberration of 42 mλ.

C. Image Post-Processing

In this work, no use is made of the on-chip processing capabilities of the sensor, but instead raw measurement data is written to a digital file for further processing in MATLAB. The four-dimensional image dataset $I[l, p, c, s]$ depends on a first index $l \in [1, N_l]$ for the "scan" direction with size N_l equal to the number of scanned lines, a second index $p \in [1, N_p]$ for the field direction, with $N_p = 4096$ pixels, a third index $c \in [1, N_c]$ representing the color channels, which has $N_c = 3$ for RGB imaging, and a fourth index $s \in [1, N_s]$, enumerating the $N_s = 8$ layers. Lines and channels are acquired line after line, and for every line N_c channels are acquired sequentially. The connotation of the term "line" is therefore temporal in the current context and must not be confused with a physical row of pixels on the sensor.

Three pre-processing steps are taken for all datasets. First, the sensor offset is corrected, which consists of two terms. The sensor has a line-to-line varying offset $I_d[l, c, s]$ which is provided in the raw data of each scan. The second offset is a lateral offset $I_l[p, c, s]$, which must be measured only once for every particular configuration of exposure time, gain, and chosen sensorlets. The offset was measured by creating a "dark" scan without illumination. The second pre-processing step is the registration that is required to compensate for the spatial offset in the scan direction of the different sensorlets used to scan the focus layers. Additionally, small registration errors arising from misalignment of the sensor tilt and rotation typically arise. This is compensated by translating the image data d_l pixels in the scan direction and d_p pixels in the field direction, given by

$$d_l[s] = \left[\frac{N_t \Delta_s}{\Delta_l}\left(s - \frac{N_s + 1}{2}\right) + \left(r - \frac{N_r + 1}{2}\right)\right] D(1 + \delta_l), \tag{14}$$

$$d_p[s] = \left[\frac{N_t \Delta_s}{\Delta_l}\left(s - \frac{N_s + 1}{2}\right) + \left(r - \frac{N_r + 1}{2}\right)\right] \delta_l, \tag{15}$$

where a constant sensorlet interval N_t is assumed, e.g., $N_t = 4$ when sensorlets 50, 54,...,78 are used. D is the number of stage steps of size Δ_t it takes to translate the image to the next row on

the sensor, given by $D = \Delta_l \cos(\beta)/M\Delta_t$. The stage step size Δ_t is chosen such that $D = 1$ for the 20× objective lens and $D = 2$ for the 10× objective lens. The value r is the row within the sensorlet at which the data is acquired, given by

$$r[s] = \begin{cases} s, & s \le 4, \\ s - 4, & s \ge 5. \end{cases} \tag{16}$$

The overall displacement is minimized by correcting s and r for the number of used sensorlets $N_s = 8$ and the number of rows in a sensorlet $N_r = 4$. The residual alignment errors δ_l and δ_p are found by scanning a checkerboard resolution target and searching the value that optimizes the correlation between the layers. This is implemented by a coarse 5×5 grid search around zero with a step size of 0.02, followed by a parabolic interpolation of the result. Finally, the image is scaled in the scan direction with a factor $\Delta_p/\Delta_t M$, to have isotropic sampling. The system is designed to have this factor equal to one for the 20× objective lens such that this scaling can be omitted in that particular case. The third pre-processing step is to correct for non-uniformity of the illumination by applying flat fielding. The illumination intensity $I_f[l, c, s]$ is found by scanning a transparent area of a resolution target and average the result over the scan direction.

The quantitative phase tomography algorithm appears to be sensitive to inhomogeneities between the focal layers originating from incomplete flat fielding and a small line-to-line instability of the readout gain. A simple mitigation is to correct pixel values for line-to-line and pixel-to-pixel variations by the average line and pixel value:

$$I'[l, p, c, z] = I[l, p, c, z] - \frac{1}{N_p}\sum_p I[l, p, c, z] - \frac{1}{N_l}\sum_l I[l, p, c, z]. \tag{17}$$

This can work because the DC offset is not important for the phase tomography algorithm.

3. SINGLE-SCAN MULTI-FOCAL VOLUMETRIC IMAGING OF THICK TISSUE LAYERS

A. Axial Deconvolution

Widefield microscopy has no optical sectioning capability, i.e., multi-focal volumetric images have optical cross talk, adding blurred structures in out-of-focus layers to the in-focus image. This can be partly compensated by the use of deconvolution techniques, in the case of sufficiently high SNR [44–46]. A drawback of applying existing 3D deconvolution approaches to multi-focal whole-slide images is the computational complexity, which scales unfavorably with lateral image size. This may be attributed to the lateral sharpening, that is a secondary goal of 3D deconvolution methods. The idea of this lateral sharpening is to partially undo the blurring by the microscope point spread function (PSF). Here, a "z-only" approach is followed, in which the goal of lateral sharpening is abandoned, and the focus is entirely on the goal of reducing the axial cross talk. This reduces the computational complexity in such a way that the computational costs are scalable with lateral image size in a straightforward way. Deconvolution would then become suitable for real-time processing with the use of parallel processing.

The deconvolution algorithm is a variant of the iterative constrained Tikhonov–Miller (ICTM) deconvolution method [47], which is based on the minimization of the function

$$\epsilon = \frac{1}{2}|p - Gd|^2 + \frac{1}{2}w|Cd|^2, \tag{18}$$

under the constraint $d \ge 0$. Here p represents the pre-processed image data vector, d is the to-be-deconvolved image data vector, G is the blurring matrix (PSF), w is the regularization parameter, and C is the regularization matrix. The constraint $d \ge 0$ is taken into account by a mapping $d = c(s)$ from the domain $-\infty < s < \infty$ to $0 \le d < \infty$. Verveer and Jovin [47] use the mapping $c(s) = s$ for $s \ge 0$ and $c(s) = 0$ for $s < 0$. The conjugate gradient method is used in [47] to minimize Eq. (18) with the aid of this mapping. An alternative mapping that could possibly work is $c(s) = \exp(s)$. The gradient vector of the error function w.r.t. d has components:

$$r_i = \frac{\partial \epsilon}{\partial d_i} = \sum_j A_{ij} d_j - b_i, \tag{19}$$

where i and j are indices labeling the vector and matrix components, and with the vector $b = G^T p$ and the matrix $A = G^T G + wC^T C$. The Hessian is

$$\frac{\partial^2 \epsilon}{\partial d_i \partial d_j} = A_{ij}. \tag{20}$$

For the unconstrained vector s we find a gradient,

$$u_i = \frac{\partial \epsilon}{\partial s_i} = c'(s_i)\, r_i \tag{21}$$

and a Hessian,

$$B_{ij} = \frac{\partial^2 \epsilon}{\partial s_i \partial s_j} = c'(s_i)\, c'(s_j)\, A_{ij} + c''(s_i)\, r_i \delta_{ij}, \tag{22}$$

where $c'(s)$ and $c''(s)$ are the first- and second-order derivative of the mapping function. For the mapping function of [47] $c'(s) = 1$ for $s > 0$ and $c'(s) = 0$ for $s \le 0$, and $c''(s) = 0$ (the delta-peak at $s = 0$ is ignored). This boils down to $u = Pr$ and $B = P^T AP$ with P a diagonal (projection) matrix with entries $P_{ii} = 1$ for $s_i > 0$ and $P_{ii} = 0$ for $s_i \le 0$. An iterative update is made of the unconstrained deconvolved image vector s at iteration step k in the search direction v following

$$s^{k+1} = s^k + \alpha v, \tag{23}$$

with step size

$$\alpha = -\frac{u^T v}{v^T B v}. \tag{24}$$

The search direction can be computed using, e.g., the conjugate gradient method or the steepest descent method. For the sake of simplicity we use the gradient in the original d-domain, i.e., $v = r$. The constrained deconvolved image vector d at iteration $k + 1$ is then found by $d^{k+1} = c(s^{k+1})$. A fixed number of

iterations N_i is used, because this removes the need for convergence testing during the optimization, which greatly contributes to the computational efficiency. For $N_i = 30$, the deconvolved images are converged to the final image with an error below the eight bit precision level. The ICTM method is applied to do only an axial deconvolution. This "z-only" approach uses an effective blurring matrix:

$$G_{ij} = \alpha_j \left[1 - \beta + \beta \mathrm{sinc}\left(\pi \frac{z_i - z_j}{\Delta z} \right)^2 \right], \qquad (25)$$

of size $N_s \times N_s$ with N_s the number of layers, $\mathrm{sinc}(t) = \sin(t)/t$, and z is the axial position in object space. The constant term represents the axial response for a uniform object (delta-peak in Fourier space), the sinc term represents the axial response for a point object (uniform in Fourier space). The weighting coefficient β takes the relative magnitude of both contributions into account. In practice, a value close to but not equal to one gives satisfactory results. In the following $\beta = 0.95$ has been used. The parameter Δz is a measure for the depth of focus for which we use $\Delta z = 19.9$ µm. The parameter α_j is used to normalize every row of the blurring matrix G such that $\sum_j G_{ij} = 1$. This ensures that G leaves the DC component unchanged, and it prevents the deconvolution to underestimate the value of d at the upper and lower scanned layers.

Regularization is needed in order to avoid division by zero if the blurring matrix is singular and/or if prior knowledge of the imaged object is available. For the axial deconvolution with the effective blurring matrix Eq. (25) this appears to be relatively unimportant. In the following $C = I$, the identity matrix, is used and $w = 10^{-6}$.

The axial deconvolution method is implemented on a GPU using MATLAB's Parallel Processing Toolbox. Deconvolving an image segment of $512 \times 4096 \times 3 \times 8$ requires 17 ms on an Nvidia Tesla P100-PCIE-16 GB GPU. The current calculation time implies a throughput of about 3 GPx/s, far exceeding the acquisition speed.

B. Scan Setup and Samples

The 10× objective lens is used to cover a large axial range for multi-focal volumetric imaging of thick slides. A sensorlet pitch $N_t = 18$ is used such that the layers have a pitch of $N_t \Delta_{ax} = 11.7$ µm and cover a total axial range of $8 N_t \Delta_{ax} = 93.5$ µm.

Two samples are used to demonstrate the imaging of thick slides. The first slide contains a 60 µm thick stage 3 human rectum cancer section, stained with hematoxylin and eosin (H&E). The second slide contains a 60 µm thick human prostate section, stained with H&E and cleared using benzyl alcohol benzyl benzoate [48].

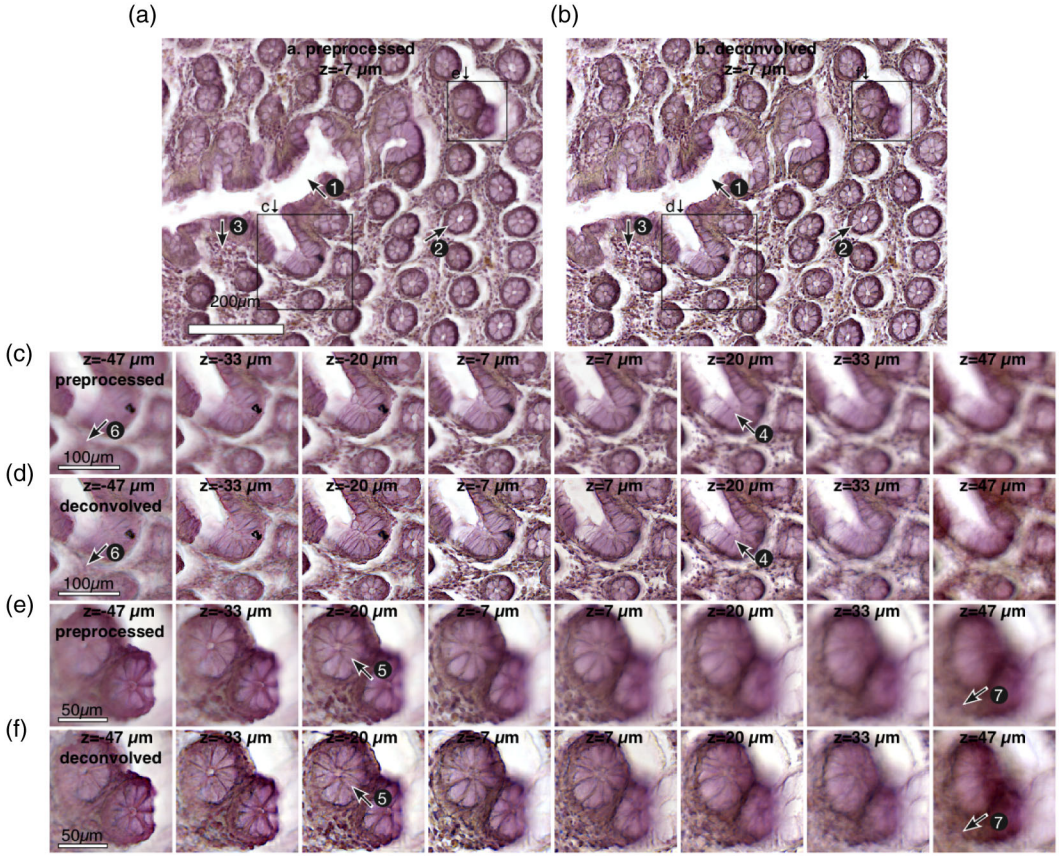

Fig. 4. Multi-layer scan of a 60 µm thick, H&E stained normal human rectal mucosa section. (a) Center layer after pre-processing. (b) Center layer after deconvolution. Visualization 1 provides the whole focal volume. (c), (e) All focal layers of a detail of the pre-processed image. (d), (f) All focal layers of a detail of the deconvolved image.

C. Results

Figure 4 shows the result of a multi-focal scan of a normal human rectum tissue section. It provides a side-by-side comparison of the image after pre-processing and the image after deconvolution. In Figs. 4(a) and 4(b), the center layer of the focal stack is shown, displaying a 800 × 800 μm^2 area of 60 μm thick mucosa tissue. Transverse sectioned colonic crypts are shown (tube-shaped mucosal structures) extended toward the bowel lumen. An edge toward the lumen is shown in the left half of the image; see Arrow 1 in Fig. 4. For two structures, a detailed view of all focal layers is provided in Figs. 4(c)–4(f).

The deconvolution algorithm clearly improves the image contrast by using the information of the neighboring layers. A higher contrast between the dark ring of cells along the crypts and the surrounding area is observed; see, e.g., Arrow 2 in Figs. 4(a) and 4(b). Also, the nuclei in the intermediate tissue show enhanced contrast; see, e.g., Arrow 3 in Figs. 4(a) and 4(b). The deconvolution reveals multiple cellular layers of the cellular inner structure of the crypt that are hard to distinguish in the pre-processed data; see, e.g., Arrow 4 in Figs. 4(c) and 4(d) and Arrow 5 in Figs. 4(e) and 4(f). Some ringing artifacts are observed in the out-of-focus layers, in particular for $z = 47$ μm; see, e.g., Arrow 6 in Figs. 4(c) and 4(d) and Arrow 7 in Figs. 4(e) and 4(f), but do not seem to harm the overall image quality too much.

The result of a multi-focal scan of a cleared prostate section is provided in Fig. 5. Again, a side-by-side comparison is provided between the pre-processed data and the deconvolved data. Figures 5(a) and 5(b) depict an 800 × 800 μm^2 area of 60 μm thick prostate tissue, showing stroma separated from prostate glandules (containing two types of cells in normal prostate). In Figs. 5(c) and 5(d) a detailed view is provided of an almost fully cleared tissue, where mainly the nuclei have remained visible. Figures 5(e) and 5(f) provide a detailed view of the differences in morphology.

The deconvolved images show an improved contrast, especially on the smaller length scales. For example, the nuclei show a good separation; see, e.g., Arrow 1 in Figs. 5(c) and 5(d). Also there is a clear separation of basal and luminal cells throughout the focal volume; see, e.g., Arrow 2 in Figs. 5(e) and 5(f). Bright circular shapes with a smooth gray outline are visible in Fig. 5(d) for $z \leq -33$ μm; see, for example, Arrow 3 in Fig. 5(d). These structures are out-of-focus nuclei of which only the center is suppressed by the axial deconvolution. This is interpreted as an artifact of the "z-only approach" to deconvolution.

4. QUANTITATIVE PHASE TOMOGRAPHY OF UNSTAINED TISSUE LAYERS

A. Quantitative Phase Tomography

We have implemented 3D phase retrieval using the quantitative phase tomography (QPT) method that was recently introduced by Descloux *et al.* [40]. This method enables the retrieval of a local phase in 3D from a stack of bright-field images by a simple

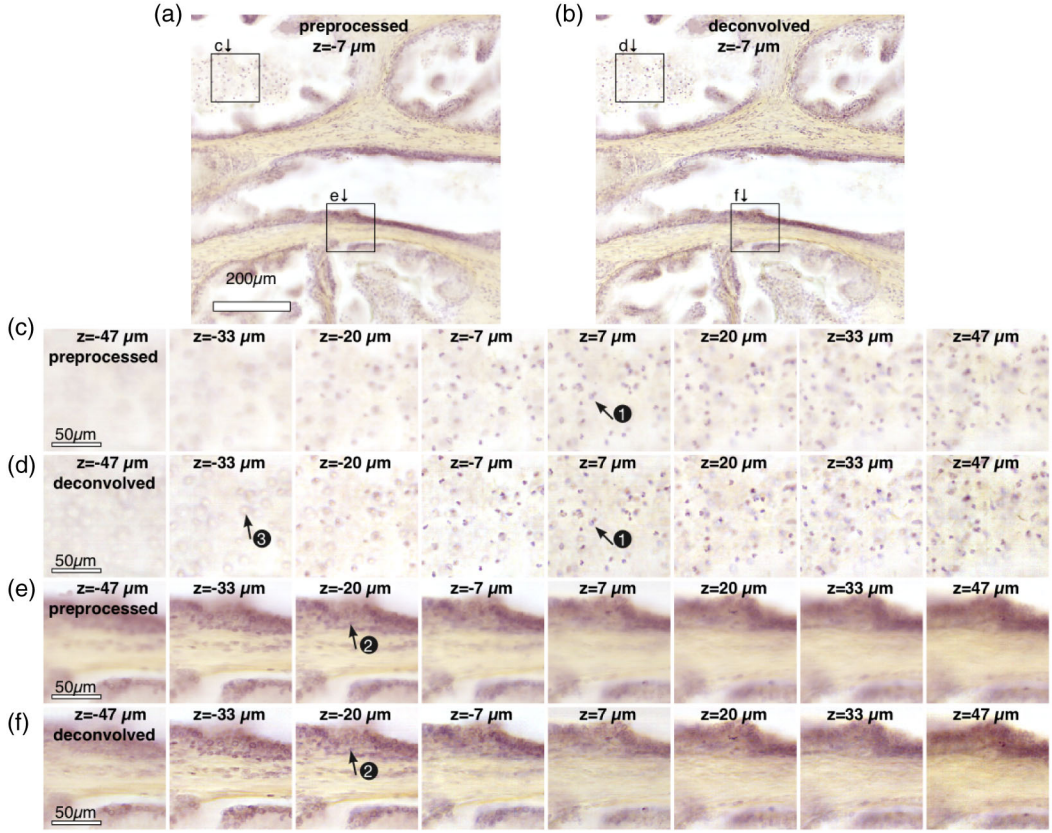

Fig. 5. Multi-layer scan of a 60 μm thick, cleared, and H&E stained human prostate section. (a) Center layer after pre-processing. (b) Center layer after deconvolution. Visualization 2 provides a side-by-side comparison of the whole focal volume. (c), (e) All focal layers of a detail of the pre-processed image. (d), (f) All focal layers of a detail of the deconvolved image.

filtering operation. Their optical analysis is based on the Born approximation, in which the 3D intensity of a weak scattering object can be written in 3D Fourier space as a function of the 3D spatial frequency vector $\vec{f}$ as

$$\hat{I}(\vec{f}) \approx I_0\delta(\vec{f}) + \hat{\Gamma}(\vec{f}) + \hat{\Gamma}^*(\vec{f}). \tag{26}$$

The first term represents a DC offset with amplitude I_0, giving rise to the delta-function $\delta(\vec{f})$ in reciprocal space. The second and third terms are the complex valued so-called cross-spectral density $\hat{\Gamma}(\vec{f})$ and its complex conjugate.

After a 3D Fourier transform from spatial frequency to real space, the cross-spectral density is related to the local phase by

$$\phi(\vec{r}) = \tan^{-1}\left(\frac{\alpha \mathrm{Im}(\Gamma(\vec{r}))}{I_0 + \alpha \mathrm{Re}(\Gamma(\vec{r}))}\right), \tag{27}$$

where α is a calibration factor and $\vec{r} = (x, y, z)$ is the spatial position vector.

Descloux *et al.* provide a description of the spatial frequency support of the cross-spectral density, which is given by the difference between any possible spatial frequency vector in the cone of directions of incidence on the illumination side and in the cone of directions of scattering on the detection side:

$$\vec{f} = \frac{n}{\lambda}\left(\begin{bmatrix}\sin\theta_i\\ \cos\theta_i\end{bmatrix} - \begin{bmatrix}\sin\theta_d\\ \cos\theta_d\end{bmatrix}\right), \quad \text{with } |\theta_i| < \arcsin\frac{\mathrm{NA}_i}{n}, \quad \text{and } |\theta_d| < \arcsin\frac{\mathrm{NA}_d}{n}, \tag{28}$$

where n is the average object refractive index, NA_i is the illumination numerical aperture, NA_d is the detection numerical aperture, and λ is the wavelength. Only the spatial frequency component along the radial direction ($f_\perp$) and along the optical axis (f_z) are given, as the optical system is rotationally symmetric around the optical (z) axis. The frequency support of the complex conjugate $\hat{\Gamma}^*$ has the same shape, but is mirrored in the $f_\perp$ axis (substitution $f_z \to -f_z$). Figure 6(a) shows the frequency support in the incoherent limit, i.e., the detection NA_d is equal to the illumination NA_i, where $\hat{\Gamma}$ fully overlaps with its complex conjugate. In contrast, in a partially coherent system, i.e., $\mathrm{NA}_i < \mathrm{NA}_d$, $\hat{\Gamma}$ can be largely separated from $\hat{\Gamma}^*$; see Fig. 6(b). The frequency support of $\hat{\Gamma}$ has an upper bound in f_z given by

$$f_m = \frac{n}{\lambda}\left(1 - \sqrt{1 - \frac{\mathrm{NA}_d^2}{n^2}}\right). \tag{29}$$

The upper bound in f_z of the complex conjugate Γ^* is

$$f_c = \frac{n}{\lambda}\left(1 - \sqrt{1 - \frac{\mathrm{NA}_i^2}{n^2}}\right). \tag{30}$$

Both f_m and f_c are indicated in Fig. 6(b). Here it can be seen that a substantial part of $\hat{\Gamma}$ can be retrieved from a stack of intensity data by using a single sideband high-pass filter K:

$$\hat{\Gamma}_+(\vec{f}) = \hat{I}(\vec{f})\hat{K}(\vec{f}),\ \hat{K}(\vec{f}) = \begin{cases}1, & f_z > f_c\\ 0, & \text{otherwise}\end{cases}. \tag{31}$$

It can be seen from Eq. 28 that, provided that the NA_i is substantially smaller than NA_d, the support of Γ is bound by the arc $(f_\perp\lambda/n + \sin\theta_i)^2 + (f_z\lambda/n - \cos\theta_i)^2 = 1$. The intersection of this curve with $f_z = f_c$ provides the smallest lateral spatial frequency f_l for which $\hat{\Gamma}_+$ is non-zero. This smallest lateral spatial frequency f_l is given by

$$f_l = \frac{n}{\lambda}\left[2\sqrt{1 - \frac{\mathrm{NA}_i^2}{n^2} - \left(1 - \frac{\mathrm{NA}_i^2}{n^2}\right)} - \frac{\mathrm{NA}_i}{n}\right]. \tag{32}$$

The highest lateral spatial frequency for which $\hat{\Gamma}_+$ is non-zero is given by the regular partially coherent cutoff frequency:

$$f_u = \frac{\mathrm{NA}_d}{\lambda} + \frac{\mathrm{NA}_i}{\lambda}. \tag{33}$$

In Fig. 6(b) both f_l and f_u are indicated.

It may be concluded that, although we do not apply explicit filtering in the lateral direction, the optical transfer function of this new phase imaging modality will be a bandpass filter in the lateral plane. The absence of low spatial frequencies is not uncommon to phase reconstructions obtained from through-focus image stacks. Methods based on solving the TIE [34–37] need to invert the Laplacian in the lateral coordinates. The Laplacian has a transfer function that depends quadratically on the (lateral) spatial frequency components, i.e., it has (near) zero transfer at low spatial frequencies. The bandpass character of the QPT modality under consideration will result in phase reconstructions in which there is considerable edge ringing. Two

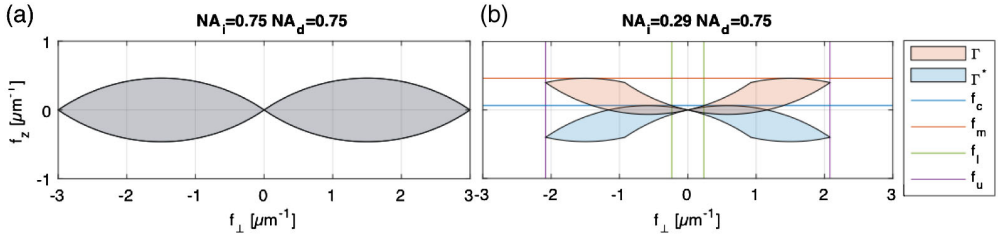

Fig. 6. Frequency support of the cross-spectral density $\hat{\Gamma}$ and its complex conjugate $\hat{\Gamma}^*$ as a function of the lateral spatial frequency $f_\perp = \sqrt{{f_x}^2 + {f_y}^2}$ and the axial spatial frequency f_z for a wavelength $\lambda = 500$nm. (a) In the incoherent limit $\mathrm{NA}_i = \mathrm{NA}_d$, the support of $\hat{\Gamma}$ fully overlaps with the support of $\hat{\Gamma}^*$. (b) Under partial coherence conditions $\mathrm{NA}_i < \mathrm{NA}_d$, $\hat{\Gamma}$ can be retrieved in the region $f_z > f_c$ and $f_l < |f_\perp| < f_u$ from the intensity data.

uniform regions separated by a sharp phase step are recognizable as flat regions with the same phase value, the border between the two regions separated by a single oscillation. A small point-like phase object is imaged as a λ/NA_d sized spot with considerable circular fringes. In fact the integral over the entire PSF must be zero, as the transfer function at zero spatial frequency is zero.

In the following we present an analysis of different trade-offs between and interdependencies of the optical system parameters. There are six system parameters for quantitative phase imaging: the NA of the detection NA_d, the NA of the illumination NA_i, the average sample refractive index n, the wavelength λ, the axial sampling distance Δz, and the number of layers scanned N_s. These parameters, however, are not independent. First, Nyquist sampling in the axial direction requires an axial sampling:

$$\Delta z = \frac{1}{2f_m} = \frac{\lambda}{2n}\left(1 - \sqrt{1 - \frac{\mathrm{NA}_d^2}{n^2}}\right)^{-1}, \tag{34}$$

where we used the relation between f_m and NA_d given in Eq. (29). Second, it appears that there is an optimal choice for the illumination NA_i given the number of scanned layers N_s. Equation (30) implies that the illumination NA_i can be derived from the lower axial cutoff frequency f_c. Generally, a lower f_c is favorable, but f_c is limited by the sampling density in the Fourier domain. For that reason f_c is selected to be half of the smallest resolvable frequency:

$$f_c = \frac{1}{2N_s\Delta z}. \tag{35}$$

The corresponding NA_i can be found by solving Eq. (30) and substituting Eq. (35) to obtain

$$\mathrm{NA}_i = n\sqrt{1 - \left(1 - \frac{1 - \sqrt{1 - \mathrm{NA}_d^2/n^2}}{N_s}\right)^2}. \tag{36}$$

Substitution in Eq. (30) leads to

$$f_c = \frac{f_m}{N_s}, \tag{37}$$

i.e., the ratio to the upper and lower axial cutoff frequencies must be equal to the number of scanned layers. It is noted that the system has no direct dependence on n, but only on the ratio's NA_d/n and NA_i/n. The current analysis shows that instead of six, there are only three degrees of freedom in the design of the optical system. A convenient set of three independent system parameters are formed by the diffraction length scale λ/NA_d, the scaled objective lens NA_d/n, and the number of scanned layers N_s.

In Fig. 7 different relevant optical system parameters are plotted as a function of N_s for four different NA_d/n values corresponding to objective lenses with $\mathrm{NA}_d = 0.2, 0.45, .75, 1.2$ and an average sample refractive index $n = 1.33$. Figure 7(a) shows the axial spatial frequency support of $\hat{\Gamma}_+$, i.e., the range of spatial frequencies between f_c and f_m. The maximum axial spatial frequency f_m is independent of the number of layers scanned but shows a strong dependence on NA_d/n. This implies that the axial resolution depends on NA comparable to conventional brightfield or fluorescence microscopes. The lower bound of the spatial frequency support f_c appears to be inversely proportional to the number of scanned layers N_s according to Eq. (37). Imaging objects with large axial size, i.e., small axial spatial frequencies, is thus realized primarily by scanning more layers.

The lateral frequency support of $\hat{\Gamma}_+$, i.e., the range of frequencies between the lower and upper cutoff spatial frequencies f_l and f_u is shown in Fig. 7(b). The lateral frequency support depends only slightly on NA_d/n. For increasing values of N_s, the lateral upper bound f_u decreases to an asymptotic value λ/NA_d. The lateral lower bound f_l decreases typically as $1/\sqrt{N_s}$, except for the practically not so relevant case $N_s = 2$ or 3 in combination with a high value of NA_d/n. Imaging objects with large lateral size, i.e., small lateral spatial frequencies, can thus be improved by scanning more layers, just as for the axial case, although the rate of improvement scales less favorably for the lateral case.

Figure 7(c) shows the partial coherence factor $\mathrm{NA}_i/\mathrm{NA}_d$, which turns out to depend hardly on NA_d/n. The partial coherence factor decreases with N_s typically as $1/\sqrt{N_s}$.

B. Numerical Implementation

The QPT method takes the following steps [40]: 1. Padding the intensity data $I(x, y, z)$ with a mirrored copy $I(x, y, -z)$ to minimize Fourier streaking occurring because of the boundary discontinuity in the axial direction. 2. Take a 3D Fourier

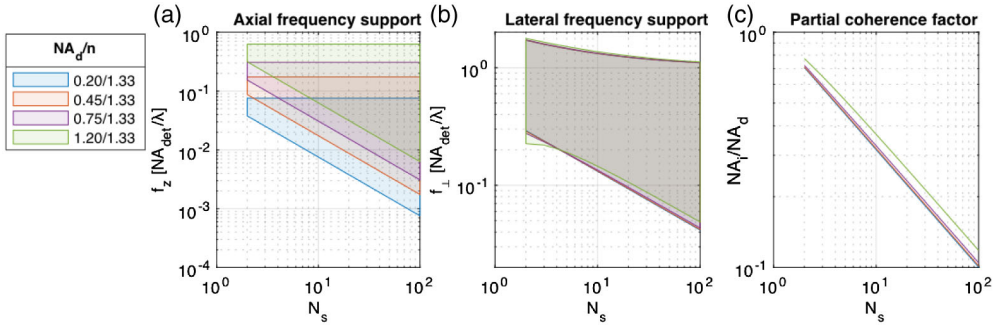

Fig. 7. Behavior of a QPT optical system expressed in normalized coordinates, as a function of the dimensionless numbers N_s and NA_d/n. (a) Axial frequency support of $\hat{\Gamma}_+$. (b) Lateral frequency support $\hat{\Gamma}_+$. (c) Partial coherence factor.

transform of the data. 3. Apply the axial spatial frequency filter. 4. Optionally, noise can be reduced by suppressing all spatial frequencies where no signal is expected based on the known spatial frequency support given in Eq. (28). 5. Take the inverse 3D Fourier transform. 6. Remove the padded data. 7. Calculate the phase using Eq. (27).

In our application, scalability to large image datasets is of utmost importance. We therefore do not implement the optional fourth step. This makes the implementation a z-only problem, just as for the deconvolution approach we propose, easily compatible with parallel processing solutions. It turns out that it is numerically convenient to implement steps 1–3, 5, and 6 using a single matrix transform:

$$\Gamma_+ = P\,I(\vec{r}) = \left[C F_z^{-1} K F_z M\right] I(\vec{r}), \tag{38}$$

where P is a $N_s \times N_s$ matrix with N_s the number of layers in the intensity stack $I(\vec{r})$. P is composed of four matrices. First, M is a $2N_s \times N_s$ matrix that pads the data using $M_{ij} = \delta_{ij} + \delta_{2N_s - i + 1, j}$, where δ is the Kronecker delta. The second matrix F has size $2N_s \times 2N_s$ and implements the Fourier transform along the z axis. Matrix K is a $2N_s \times 2N_s$ diagonal matrix implementing the spatial frequency filter. The data is inverse Fourier transformed using the $2N_s \times 2N_s$ matrix F^{-1}. Finally, the mirrored data is removed by the $N_s \times 2N_s$ matrix $C_{ij} = \delta_{ij}$ for $i, j \leq N_s$ and zero otherwise. The matrix P has to be calculated only once and can then be applied to every coordinate (x, y) independently. This method is particularly suitable for small numbers of layers, where the need for extra memory allocation that is associated with data mirroring outweighs the high efficiency of the fast Fourier transform compared to a matrix multiplication. This method is implemented for execution on a GPU using the MATLAB Parallel Processing Toolbox. Calculating the phase of a $2048 \times 2048 \times 8$ image requires about 50 ms using an Nvidia Tesla P100-PCIE-16 GB. This implies a throughput of 670 MPx.

C. Scan Setup and Samples

The QPT method is implemented based on scans with the 20× objective lens, the green color channel, and the use of $N_s = 8$ scan lines simultaneously. Given that $\mathrm{NA}_d = 0.75$, and assuming a sample refractive index of $n = 1.33$ and a wavelength of $\lambda = 500\mathrm{nm}$, Eq. (34) results in a target Δz of 1.05 μm. To guarantee Nyquist sampling a sensorlet interval $N_t = 6$ is chosen, leading to $\Delta z = 0.93$ μm. This gives a total covered axial range $N_s \Delta z = 7.4$ μm. Now Eq. (35) provides $f_c = 1/14.9$ μm, and solving Eq. (30) gives $\mathrm{NA}_i = 0.29$. As a result, the system will have a lateral resolution of $1/f_u = 0.48$ μm, as follows from Eq. (33), it images structures with a lateral size up to $1/f_l = 4.3$ μm, according to Eq. (32), and Eq. (29) provides a maximum axial spatial frequency $f_m = 1/2.2$ μm.

Three samples are used to demonstrate the QPT method. The first slide contains a 5 μm thick human prostate tissue microarray (TMA) section labeled with Kreatech ERBB2 (17q12)/SE 17 FISH probe (product number KBI-10701) for the detection of amplification of the ERBB2 (also known as HER-2/*neu*) gene via fluorescence *in situ* hybridization (FISH) [49]. The second slide is a 4 μm thick stage 3 human rectum cancer section with immunofluorescence labeling. Three proteins are labeled: Desmin (IgG1 M antibody labeled with Alexa Fluor 488), which is highly expressed in muscle cells; CD31 (IgG R antibody labeled with Alexa Fluor 546), which is a is a marker for blood vessels; and D2-40 (labeled with Alexa Fluor 594), which is used as a marker of lymphatic endothelium. Additionally, the slide is stained with DAPI, labeling the nuclei. The third slide is an 4 μm thick human prostate section. The slide is deparaffinized and embedded in xylene but not further processed for staining.

D. Results

Figure 8 shows the result for the ERBB2 slide. Figure 8(a) shows all eight layers of the pre-processed image data. A detail is selected in which a single cell is visible. The local phase computed from this data is shown in Fig. 8(b). A diverging colormap is used to display the phase values [51], with blue corresponding to negative values, black to zero, and green to positive values. The total estimated optical thickness of the sample is obtained by summing the local phase over all layers. The result is shown over three length scales, zooming in with a factor of 64: Fig. 8(c) has the full width of a single scan lane, Fig. 8(d) displays an intermediate length scale, and Fig. 8(e) corresponds to the detail shown in Figs. 8(a) and 8(b). Figures 8(f)–8(h) show a color-coded maximum intensity projection of the same area. The pixel values in this image have an intensity corresponding to the maximum phase along the axial direction and a color corresponding to the depth at which the maximum was found. The used colormap, shown at the left side of Fig. 8(f), has a uniform luminescence and an equidistant color spacing. This minimizes the visual cross talk between depth and intensity. As a reference, Figs. 8(i)–8(k) provide a maximum intensity projection of a multi-focal fluorescence image of the same section, for the same areas. This fluorescence image was obtained in previous research [26] using a multi-focal multi-line confocal scanner. A rigid transform is used to register the fluorescence reference image to the phase image, where the optimal transform is found by minimizing the root mean square distance between a series of manually selected landmarks.

In the pre-processed images shown in Fig. 8(a) the outline of the cells are visible at low contrast, as well as a few spots that change from bright to dark through focus; see Arrows 1 and 2 in Fig. 8(a). The corresponding local phase in Fig. 8(b) shows the contours of the cells and cell organelles with a high contrast. The QPT algorithm is able to reveal the axial position ($z \approx 1.7$ μm) of the sources of the bright and dark spots; see Arrow 3 in Fig. 8(c). The lateral bandpass behavior of the QPT algorithm causes ringing, which is particularly visible as a blue (negative) local phase surrounding the cell; see for example Arrow 4 in Fig. 4(b). Also in the total optical thickness, displayed in Figs. 8(c)–8(e), the cell contours and tissue structure are clearly revealed. The bandpass characteristic is again apparent, for example indicated by Arrow 5 in Fig. 8(d) and Arrow 6 in Fig. 8(e). The color-coded maximum intensity projection appears particularly suitable for imaging at larger length scales. For example, the wall of a blood vessel with a size of ∼100 μm is clearly visible; see Arrow 7 in Fig. 8(f). Comparing the results

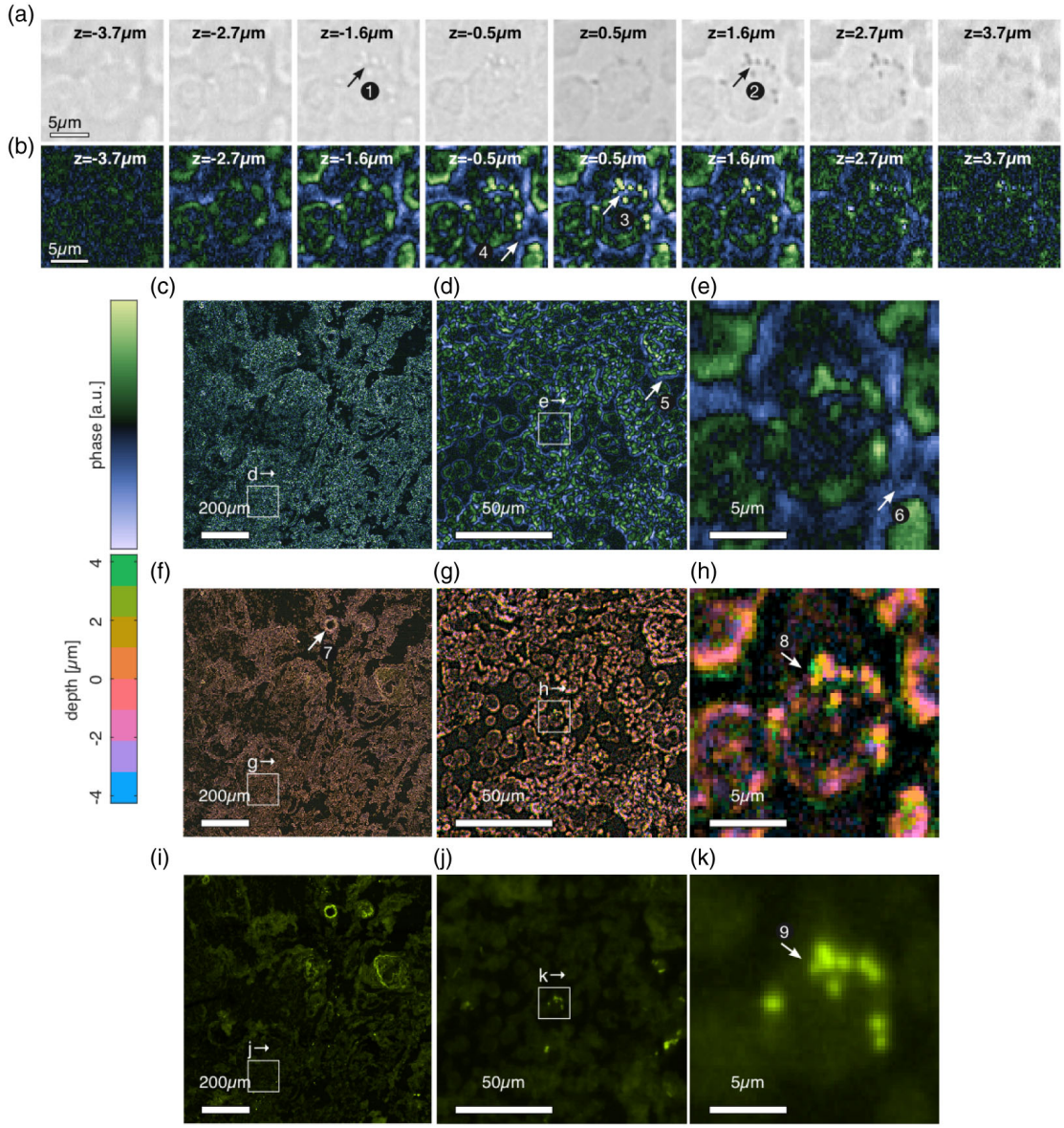

Fig. 8. Multi-layer phase contrast image of a 5 μm thick human prostate TMA section prepared for ERBB2 detection using FISH. The layers have an axial spacing of 0.93 μm spanning 7.4 μm in total. (a) Detail of the raw image data after pre-processing. (b) Retrieved phase of the corresponding area. (c)–(e) Total optical thickness, shown over three length scales. Visualization 3 shows this in more detail. (f)–(h) Color-coded maximum intensity projection of the same area; see Visualization 4. (i)–(k) Confocal fluorescence image of the same section [26] showing the same area; see Visualization 5.

with the fluorescence images, a surprisingly clear correspondence is found between the cell organelles observed in the phase images and the FISH labeled sites; compare Arrow 8 in Fig. 8(h) with Arrow 9 in Fig. 8(k).

Figure 9 shows the result for the human rectum section following the same structure as Fig. 8. The sample is shown over three length scales, zooming in with a factor of 16, ranging from the full width of a single scan lane to the same detail shown in Figs. 9(a) and 9(b). Figures 9(i)–9(k) provide a multi-color widefield fluorescence image of a directly adjacent, identically prepared slide that we reported on in earlier research [50]. This image was registered to the phase image, just as for the previous case.

The insets show the cross section of normal crypts in the human rectum. The crypt lumen indicated by Arrow 1 in Fig. 9(e) and cell walls indicated by Arrow 2 in Fig. 9(e) are clearly visible. The phase images are able to reveal relevant structure in the tissue on the larger length scales as well, in particular in the color-coded maximum intensity projection. For example, the top half of the largest zoom level shows the submucosa with two veins [see Arrow 3 in Fig. 9(f)], the muscularis mucosae [see Arrow 4 in Fig. 9(f)], while the bottom half shows the normal mucosa with the crypts, indicated by Arrow 5 in Fig. 9(f). A clear correspondence is found with the fluorescence images. For example, the red labeled microvessel indicated by Arrow 6 in Fig. 9(h) can also be observed in the phase contrast images; see

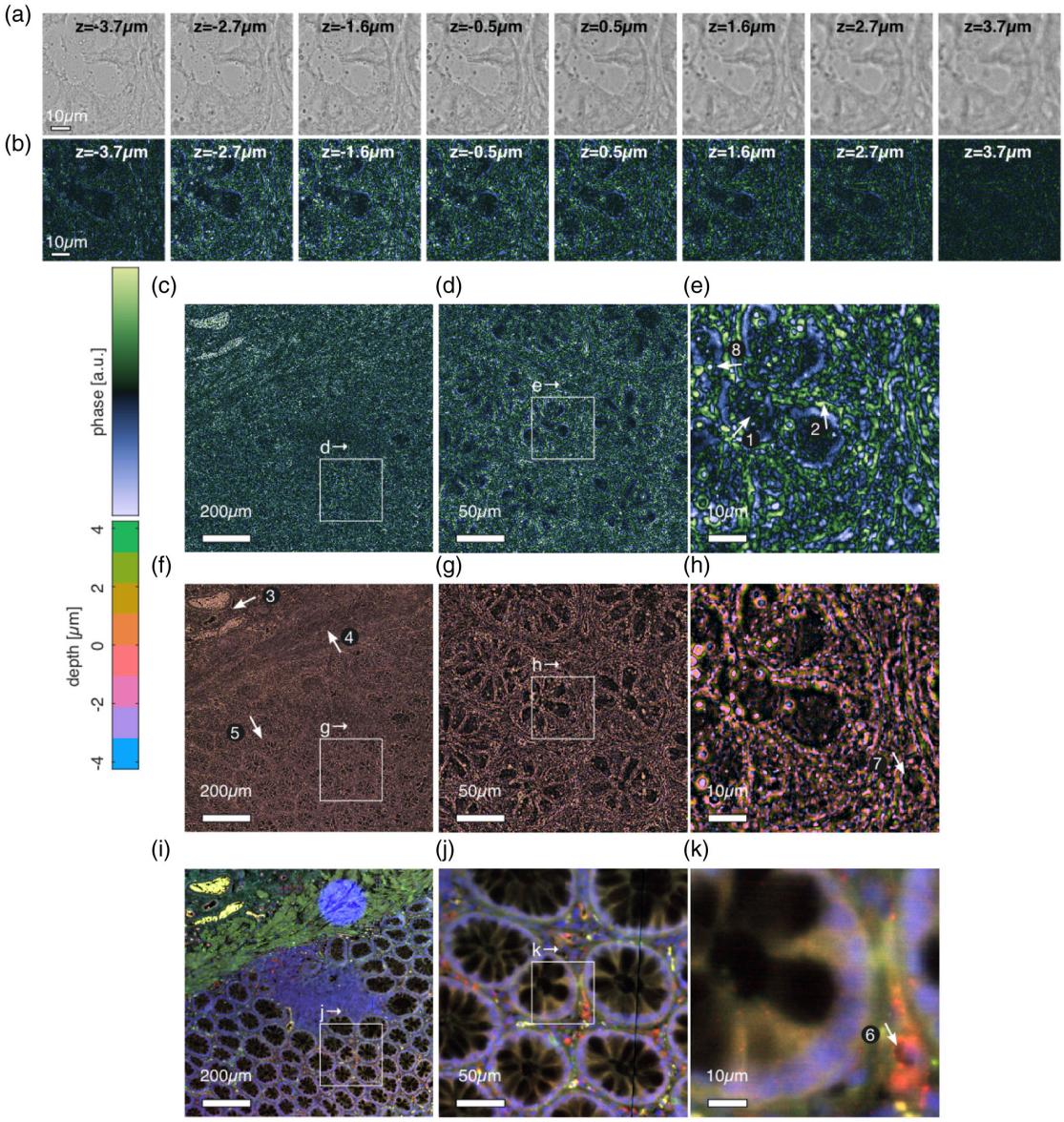

Fig. 9. Multi-layer phase contrast image of a 4 μm thick human rectum section. The layers have an axial spacing of 0.93 μm spanning 7.4 μm in total. (a) Detail of the raw image data after pre-processing. (b) Retrieved phase of the corresponding area. (c)–(e) Total optical thickness, shown over three length scales. Visualization 6 shows this in more detail. (f)–(h) Color-coded maximum intensity projection of the same area; see Visualization 7. (i)–(k) Widefield multi-color fluorescence image of an directly adjacent section [50] showing the same area; see Visualization 8.

Arrow 7 in Fig. 9(h). At the smallest zoom level small spots are visible that are not present in the fluorescence control images. Interestingly, some of them have a negative phase value [see, e.g., Arrow 8 in Fig. 9(e)], suggesting that this might be regions with a lower refractive index than the surrounding tissue structures, such as water droplets or air bubbles. Around the spots, ringing is visible in correspondence with the expected bandpass behavior.

Figure 10 shows the result for the human prostate section following the same structure as Figs. 8 and 9. The result is shown over three length scales, zooming in over a factor of 25, from the full width of a single scan lane down to a columnar epithelium layer around a lumen. As a reference an H&E stained section from the same tissue block is shown, registered to the phase image. It is noted, however, that this section was not directly adjacent to the unstained section used for phase imaging, and that therefore the overall structure is not corresponding closely to the structure of the phase images of the unstained slide.

The cell borders of the epithelium are clearly visible in the phase images; see Arrow 1 in 10(e). Also apparent are the outline of the nuclei, e.g., as indicated by Arrow 2 in Fig. 10(h) and the structure of the stroma, e.g., as indicated by Arrow 3 in Fig. 10(g). The overall features (highlighting edges and near-point like objects) are the same as for the other two cases.

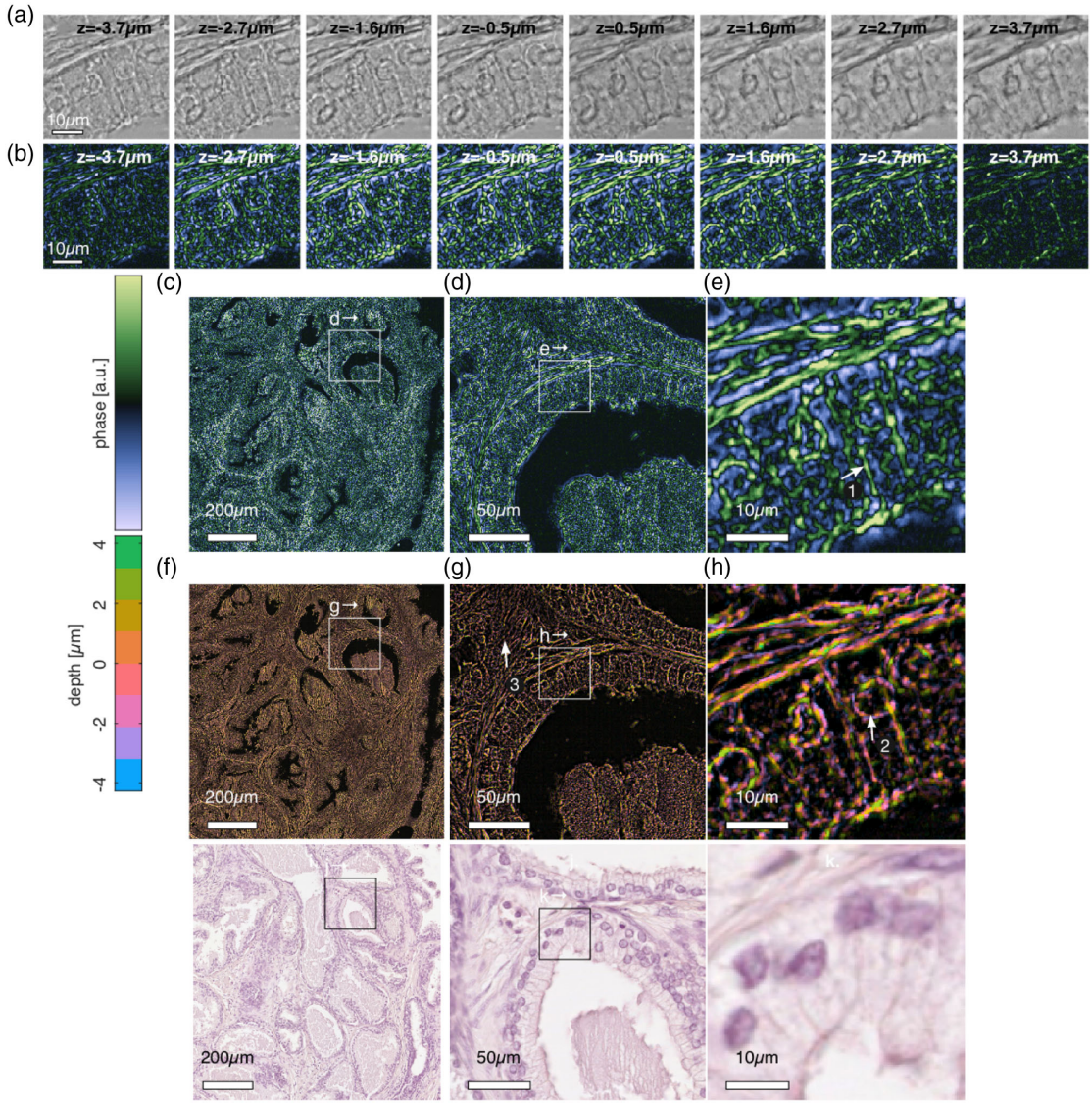

Fig. 10. Multi-layer phase contrast image of an unstained, 4 μm thick human prostate section. The layers have an axial spacing of 0.93 μm spanning 7.4 μm in total. (a) Detail of the raw image data after pre-processing. (b) Retrieved phase of the corresponding area. (c)–(e) Total optical thickness, shown over three length scales. Visualization 9 shows this in more detail. (f)–(h) Color-coded maximum intensity projection of the same area; see Visualization 10. (i)–(k) H&E stained section of the same tissue block showing the same area; see Visualization 11.

5. CONCLUSION

In conclusion, we present a WSI system based on a multi-line CMOS sensor devised by Philips [23–25]. A tilted image plane makes it possible to acquire image data from multiple focal slices simultaneously. The architecture with a single image sensor provides inherent registration of different color channels and focal layers. The "push-broom" scanning approach results in a field of view that is in principle unlimited, and a reduced need for stitching. The system achieves image acquisition with a very high throughput of ~400MPx/s, and can be used for 2D full-color imaging of absorption stained slides with continuous autofocus, but also for multi-focus imaging.

The scanner platform enables novel contrast modalities based on computational imaging approaches. Multi-focal volumetric imaging of thick pathology samples is demonstrated, where eight layers are acquired in a single scan. Samples of 60 μm thick are imaged with a resolution of 0.5 μm. Deconvolution is used to improve contrast, which is inherently low for such thick specimens. A simplified form of the ICTM deconvolution method is proposed, targeting suppression of out-of-focus light only, and ignoring lateral resolution improvement, enabling very high processing speeds of about 3 GPx/s, far exceeding the acquisition speed. This indicates that this approach is suitable for on-line, and potentially on-chip, processing.

The sectioning of the focal layers could possibly be improved upon by combining the proposed "z-only" deconvolution with a multi-scale image approach [52]. Decomposition of the image data in a wavelet representation would enable a different setting of the depth of focus parameter for every lateral length scale in the wavelet domain [53]. In this way, the deconvolution

would become effectively dependent on the lateral spatial frequency content, while at the same time avoiding a (much) larger blurring matrix.

Next, phase imaging based on the recently introduced QPT method [40] is demonstrated. A simplified form of the algorithm, based on "z-only" processing is proposed for this modality as well, giving data processing speeds of $\sim$0.67 GPx/s, exceeding the acquisition speed. A system design study of the phase imaging modality is developed in this paper, leading to a description in which the axial and lateral spatial frequency support, as well as the partial coherence factor, are entirely given by the diffraction length scale λ/NA_d, the scaled imaging NA_d/n, and the number of scanned layers N_s. The image formation theory points to an in-plane transfer function that has the character of a bandpass spatial frequency filter. The lower and upper cutoff spatial frequencies for the current setup (with objective lens $\mathrm{NA}_d = 0.75$ and condenser $\mathrm{NA}_i = 0.26$) are at $1/f_l = 4.3$ μm and $1/f_u = 0.48$ μm. This bandpass transfer results in phase images where the near-point like objects, with a size of $\sim$1 μm, are highlighted, and where phase step edges are recognizable by a single oscillation in the direction orthogonal to the edges. Comparison of QPT images of a slide prepared for FISH to the fluorescence reference image reveals that QPT imaging is able to image the sites labeled for FISH imaging. QPT imaging of an unstained tissue slide and comparison to immunofluorescence imaging shows that phase imaging can provide additional structural tissue information.

A next step for the QPT method could be to include a form of in-plane image processing, targeting to overcome the primary limitation of the QPT method of having a zero transfer function at low spatial frequencies, a limitation shared with TIE-based methods of solving the phase from through-focus image stacks. This next step should be accompanied with a careful balancing of signal reconstruction and noise amplification at low spatial frequencies, following the lines of, e.g., [36,37] for TIE-based methods. At the same time, sacrifices to computational speed should not be too large, as speed is a need for real-time processing of $\sim$cm^2 tissue areas scanned in $\sim$1 min.

Funding. Stichting voor de Technische Wetenschappen (13892).

Acknowledgment. We wish to thank Bas Hulsken, Rik Kneepkens, Jelte Peter Vink, Mathijs Rem, Gert-Jan van den Braak, and Aslihan Arslan Carisey of Philips Digital & Computational Pathology; Hein Sleddens of the Erasmus University Medical Center; and Hans Morreau of the Leiden University Medical Center for valuable research input.

Disclosures. Philips Digital & Computational Pathology has made their prototype scanner infrastructure available to us for this research (F).

REFERENCES

1. L. Pantanowitz, J. H. Sinard, W. H. Henricks, L. A. Fatheree, A. B. Carter, L. Contis, B. A. Beckwith, A. J. Evans, A. Lal, and A. V. Parwani, "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**, 1710–1722 (2013).
2. R. Randell, R. A. Ruddle, and D. Treanor, "Barriers and facilitators to the introduction of digital pathology for diagnostic work," Stud. Health Technol. Inform. **216**, 443–447 (2015).
3. D. R. J. Snead, Y.-W. Tsang, A. Meskiri, P. K. Kimani, R. Crossman, N. M. Rajpoot, E. Blessing, K. Chen, K. Gopalakrishnan, P. Matthews, N. Momtahan, S. Read-Jones, S. Sah, E. Simmons, B. Sinha, S. Suortamo, Y. Yeo, H. El Daly, and I. A. Cree, "Validation of digital pathology imaging for primary histopathological diagnosis," Histopathology **68**, 1063–1072 (2016).
4. E. Abels and L. Pantanowitz, "Current state of the regulatory trajectory for whole slide imaging devices in the USA," J. Pathol. Inform. **8**, 23 (2017).
5. P. W. Hamilton, Y. Wang, and S. J. McCullough, "Virtual microscopy and digital pathology in training and education," Acta Pathol. Microbiol. Immunol. Scand. **120**, 305–315 (2012).
6. K. Foster, "Medical education in the digital age: digital whole slide imaging as an e-learning tool," J. Pathol. Inform. **1**, 14 (2010).
7. L. Pantanowitz and A. V. Parwani, "Education," in *Digital Pathology* (Springer, 2016), pp. 71–78.
8. E. A. Krupinski, A. K. Bhattacharyya, and R. S. Weinstein, "Telepathology and digital pathology research," in *Digital Pathology* (Springer, 2016), pp. 41–54.
9. P. W. Hamilton, P. Bankhead, Y. Wang, R. Hutchinson, D. Kieran, D. G. McArt, J. James, and M. Salto-Tellez, "Digital pathology and image analysis in tissue biomarker research," Methods **70**, 59–73 (2014).
10. B. Hulsken, D. Vossen, and S. Stallinga, "High NA diffractive array illuminators and application in a multi-spot scanning microscope," J. Euro. Opt. Soc. **7**, 12026 (2012).
11. J. Wu, X. Cui, G. Zheng, Y. M. Wang, L. M. Lee, and C. Yang, "Wide field-of-view microscope based on holographic focus grid illumination," Opt. Lett. **35**, 2188–2190 (2010).
12. S. Pang, C. Han, M. Kato, P. W. Sternberg, and C. Yang, "Wide and scalable field-of-view Talbot-grid-based fluorescence microscopy," Opt. Lett. **37**, 5018–5020 (2012).
13. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**, 739–745 (2013).
14. K. Guo, S. Dong, and G. Zheng, "Fourier ptychography for brightfield, phase, darkfield, reflective, multi-slice, and fluorescence imaging," IEEE J. Sel. Top. Quantum Electron. **22**, 77–88 (2016).
15. L. Tian, Z. Liu, L.-H. Yeh, M. Chen, J. Zhong, and L. Waller, "Computational illumination for high-speed *in vitro* Fourier ptychographic microscopy," Optica **2**, 904–911 (2015).
16. F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," Annu. Rev. Pathol.: Mech. Dis. **8**, 331–359 (2013).
17. 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**, 4 (2006).
18. M. G. Rojo, G. B. García, C. P. Mateos, J. G. García, and M. C. Vicente, "Critical comparison of 31 commercially available digital slide systems in pathology," Int. J. Surg. Pathol. **14**, 285–305 (2006).
19. H. Netten, L. J. van Vliet, F. R. Boddeke, P. de Jong, and I. T. Young, "A fast scanner for fluorescence microscopy using a 2-D CCD and time delayed integration," Bioimaging **2**, 184–192 (1994).
20. M. Montalto, R. Filkins, and R. McKay, "Autofocus methods of whole slide imaging systems and the introduction of a second-generation independent dual sensor scanning method," J. Pathol. Inform. **2**, 44 (2011).
21. M. E. Bravo-Zanoguera, C. A. Laris, L. K. Nguyen, M. Oliva, and J. H. Price, "Dynamic autofocus for continuous-scanning time-delay-and-integration image acquisition in automated microscopy," J. Biomed. Opt. **12**, 034011 (2007).
22. B. Hulsken, "Autofocus based on differential measurements," U.S. patent 9,832,365 (28 November 2017).
23. B. Hulsken, "Method for simultaneous capture of image data at multiple depths of a sample," U.S. patent 9,910,258 (6 March 2018).
24. B. Hulsken, "Scanning imaging system with a novel imaging sensor with gaps for electronic circuitry," U.S. patent 10,091,445 (2 October 2018).

25. B. Hulsken and S. Stallinga, "Sensor for microscopy," U.S. patent 10,353,190 (16 July 2019).
26. L. van der Graaff, G. J. L. H. van Leenders, F. Boyaval, and S. Stallinga, "Multi-line fluorescence scanning microscope for multi-focal imaging with unlimited field of view," Biomed. Opt. Express **10**, 6313–6339 (2019).
27. S. M. Shakeri, L. J. van Vliet, and S. Stallinga, "Impact of partial coherence on the apparent optical transfer function derived from the response to amplitude edges," Appl. Opt. **56**, 3518–3530 (2017).
28. T. C. Cornish, R. E. Swapp, and K. J. Kaplan, "Whole-slide imaging: routine pathologic diagnosis," Adv. Anatomic Pathol. **19**, 152–159 (2012).
29. N. Roberts, D. Magee, Y. Song, K. Brabazon, M. Shires, D. Crellin, N. M. Orsi, R. Quirke, P. Quirke, and D. Treanor, "Toward routine use of 3D histopathology as a research tool," Am. J. Pathol. **180**, 1835–1842 (2012).
30. E. A. El-Gabry, A. V. Parwani, and L. Pantanowitz, "Whole-slide imaging: widening the scope of cytopathology," Diagnostic Histopathol. **20**, 456–461 (2014).
31. W. E. Khalbuss, L. Pantanowitz, and A. V. Parwani, "Digital imaging in cytopathology," Pathol. Res. Int. **2011**, 1–10 (2011).
32. T. Kalinski, R. Zwönitzer, S. Sel, M. Evert, T. Guenther, H. Hofmann, J. Bernarding, and A. Roessner, "Virtual 3D microscopy using multiplane whole slide images in diagnostic pathology," Am. J. Clin. Pathol. **130**, 259–264 (2008).
33. Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," Nat. Photonics **12**, 578–589 (2018).
34. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," J. Opt. Soc. Am. **73**, 1434–1441 (1983).
35. D. Paganin and K. A. Nugent, "Noninterferometric phase imaging with partially coherent light," Phys. Rev. Lett. **80**, 2586–2589 (1998).
36. L. Waller, M. Tsang, S. Ponda, S. Y. Yang, and G. Barbastathis, "Phase and amplitude imaging from noisy images by Kalman filtering," Opt. Express **19**, 2805–2815 (2011).
37. E. Bostan, E. Froustey, M. Nilchian, D. Sage, and M. Unser, "Variational phase imaging using the transport-of-intensity equation," IEEE Trans. Image Process. **25**, 807–817 (2016).
38. T. Kim, R. Zhou, M. Mir, S. D. Babacan, P. S. Carney, L. L. Goddard, and G. Popescu, "White-light diffraction tomography of unlabelled live cells," Nat. Photonics **8**, 256–263 (2014).
39. R. Horstmeyer, J. Chung, X. Ou, G. Zheng, and C. Yang, "Diffraction tomography with Fourier ptychography," Optica **3**, 827–835 (2016).
40. A. Descloux, K. S. Grußmayer, E. Bostan, T. Lukes, A. Bouwens, A. Sharipov, S. Geissbuehler, A.-L. Mahul-Mellier, H. A. Lashuel, M. Leutenegger, and T. Lasser, "Combined multi-plane phase retrieval and super-resolution optical fluctuation imaging for 4D cell microscopy," Nat. Photonics **12**, 165–172 (2018).
41. S. Abeytunge, Y. Li, B. Larson, R. Toledo-Crow, and M. Rajadhyaksha, "Rapid confocal imaging of large areas of excised tissue with strip mosaicing," J. Biomed. Opt. **16**, 050504 (2011).
42. S. M. Shakeri, B. Hulsken, L. J. van Vliet, and S. Stallinga, "Optical quality assessment of whole slide imaging systems for digital pathology," Opt. Express **23**, 1319–1336 (2015).
43. S. Stallinga, "Finite conjugate spherical aberration compensation in high numerical-aperture optical disc readout," Appl. Opt. **44**, 7307–7312 (2005).
44. G. M. P. van Kempen, L. J. van Vliet, P. J. Verveer, and H. T. M. van der Voort, "A quantitative comparison of image restoration methods for confocal microscopy," J. Microsc. **185**, 354–365 (1997).
45. R. Heintzmann, "Estimating missing information by maximum likelihood deconvolution," Micron **38**, 136–144 (2007).
46. J. B. Sibarita, "Deconvolution microscopy," Adv. Biochem. Eng./Biotechnol. **95**, 201–243 (2005).
47. P. J. Verveer and T. M. Jovin, "Acceleration of the ICTM image restoration algorithm," J. Microsc. **188**, 191–195 (1997).
48. A. Azaripour, T. Lagerweij, C. Scharfbillig, A. E. Jadczak, B. Willershausen, and C. J. Van Noorden, "A survey of clearing techniques for 3D imaging of tissues with special reference to connective tissue," Progr. Histochem. Cytochem. **51**, 9–23 (2016).
49. G. Pauletti, W. Godolphin, M. F. Press, and D. J. Slamon, "Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization," Oncogene **13**, 63–72 (1996).
50. L. van der Graaff, F. Boyaval, L. J. van Vliet, and S. Stallinga, "Fluorescence imaging for whole slide scanning using LED-based color sequential illumination," Proc. SPIE **10679**, 106790D (2018).
51. F. Crameri, "Geodynamic diagnostics, scientific visualisation and StagLab 3.0," Geosci. Model Develop. **11**, 2541–2562 (2018).
52. P. Moulin, "Multiscale image decompositions and wavelets," in *The Essential Guide to Image Processing* (Elsevier, 2009), pp. 123–142.
53. C. Vonesch and M. Unser, "A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution," IEEE Trans. Image Process. **17**, 539–549 (2008).