

## MIT Open Access Articles

*Correction of circumferential and longitudinal motion distortion in high-speed catheter/endoscope-based optical coherence tomography*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Nguyen, Tan Huu, Ahsen, Osman Oguz, Liang, Kaicheng, Zhang, Jason, Mashimo, Hiroshi et al. 2021. "Correction of circumferential and longitudinal motion distortion in high-speed catheter/endoscope-based optical coherence tomography." Biomedical Optics Express, 12 (1).

**As Published:** 10.1364/BOE.409074

**Publisher:** The Optical Society

**Persistent URL:** <https://hdl.handle.net/1721.1/143545>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.





# Correction of circumferential and longitudinal motion distortion in high-speed catheter/endoscope-based optical coherence tomography

TAN HUU NGUYEN,<sup>1,2</sup> OSMAN OGUZ AHSEN,<sup>1</sup> KAICHENG LIANG,<sup>1</sup>  ID  
JASON ZHANG,<sup>1</sup> HIROSHI MASHIMO,<sup>3,4</sup> AND JAMES G. FUJIMOTO<sup>1,\*</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup>PathAI Inc., 120 Brookline Ave, Boston, MA 02215, USA

<sup>3</sup>Veterans Affairs Boston Healthcare System, MA 02130, USA

<sup>4</sup>Harvard Medical School, MA 02130, USA

\*jgf@mit.edu

**Abstract:** Catheter/endoscope-based optical coherence tomography (OCT) is a powerful modality that visualizes structural information in luminal organs. Increases in OCT speed have reduced motion artifacts by enabling acquisition faster than or comparable to the time scales of physiological motion. However motion distortion remains a challenge because catheter/endoscope OCT imaging involves both circumferential and longitudinal scanning of tissue. This paper presents a novel image processing method to estimate and correct motion distortion in both the circumferential and longitudinal directions using a single en face image from a volumetric data set. The circumferential motion distortion is estimated and corrected using the en face image. Then longitudinal motion distortion is estimated and corrected using diversity of image features along the catheter pullback direction. Finally, the OCT volume is resampled and motion corrected. Results are presented on synthetic images and clinical OCT images of the human esophagus.

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

## 1. Introduction

Catheter/endoscope-based optical coherence tomography (OCT) is a powerful imaging modality capable of capturing depth-resolved tissue features in luminal organs. Since the development of the first catheter-based OCT probe more than 20 years ago [1], many variations of scanning probes have been developed. These scanning probes consist of micro-optics to deflect and focus light onto the tissue and beam scanning mechanisms which can acquire cross sectional or volumetric OCT data. The imaging probe can be miniaturized so that it can be introduced through instrument channels of endoscopes and image areas of interest under endoscopic guidance. Depending on the optical beam scanning direction relative to the longitudinal axis of the probe, catheter OCT probes can be divided into either forward-viewing [2–10], scanning the beam in front of the probe, or side-viewing, scanning the beam perpendicular to the longitudinal axis. Side-viewing probes are more popular for imaging luminal organs and can be further divided based on the type of scanning mechanism (proximal or distal). Proximal scanning probes [11–16] use a fiber optic rotary joint to transmit the rotation from a proximal motor to the distal end of the probe. Conversely, distal scanning probes [17–20] use micro-motors at their tip for circumferential beam scanning. The circumferential scanning is combined with a pullback retraction of the optical assembly to perform a helical scan and acquire volumetric data. In this paper, we limit our attention to motion distortion associated with side-viewing probes.

There are different types of motion distortion in side-viewing OCT probes including: radial, circumferential, and longitudinal distortion. Radial distortion is often caused by changes in the distance between the tissue lumen and scanning probe, leading to variation in the axial position of the tissue and also the spatial sampling pitch. However, in catheter/endoscopic OCT, this motion is usually small, because the probe is in contact with the tissue. The second type of distortion, circumferential distortion, is sometimes referred to as non-uniform rotational distortion (NURD) and is caused by variations in the circumferential scan position or speed. In proximal scanning systems, NURD is typically caused by the mechanical friction between the catheter torque coil and surrounding sheath [21]. It is further exacerbated by physiological motion of the tissue lumen and bending of the imaging catheter. The use of distal scanning micro-motor OCT probes dramatically reduces NURD associated with mechanical friction in the probe. NURD is reduced, but still occurs in micro-motor probes, because the motors have bearing friction and are usually driven open loop [22]. Circumferential motion distortion can be corrected by non-rigid registration using motion estimated from fiducial markers on the probe housing [22], speckle statistics between consecutive A-scans from a dynamic light scattering model [23], directly from OCT en face images by assuming there are slowly varying structures that are oriented roughly parallel to the pullback axis [24], or by maximizing the similarity between consecutive cross-sectional frames after alignment [25]. There are also micro-motor engineering solutions to reduce NURD, such as using synchronous micro-motors to reduce the speed variation with respect changing loads or temperature [26,27]. These motors have been used for intravascular OCT applications which require high rotational speed, small dimensions, and a very low maximum torque ( $\sim$ a few  $\mu\text{Nm}$ ). Conversely, micro-motors used in endoscopic OCT scan a larger diameter lumen and require large beam steering optics. This necessitates larger micro-motors which also have higher moment of inertia armatures.

Circumferential motion distortion can also be caused by lateral movement of the tissue on the time scale of the sequential circumferential scans. The data acquisition in side-viewing OCT scanning is similar to a rolling shutter in commercial complementary metal-oxide semiconductor (CMOS) cameras where sequential rows of the camera sensor are exposed at different times [28]. Therefore, there is image distortion caused by moving objects in both methods, because every row of the CMOS camera or every B-scan in OCT samples the object at a different time. For OCT, even if the rotational beam scanning is uniform, circumferential distortion can still occur if there is relative displacement between the tissue and catheter during sequential rotational scans. This relative displacement can be caused by peristalsis, respiration, or cardiac motion, as well as coupled into the catheter from an external source.

The third type of motion distortion is longitudinal distortion, which is manifest as stretched or compressed areas in en face OCT images. This distortion is often caused by pullback speed variations arising from mechanical friction of the micro-motor and optical assembly with the probe sheath or by relative longitudinal motion (sliding) between the tissue and catheter from physiological motion. This type of distortion poses a challenge to interpreting en face OCT images in clinical applications. For example, if en face OCT images are used to detect dysplasia in patients with Barrett's esophagus, longitudinal motion distortion may obscure the mucosal features or cause them to appear distorted.

Software and hardware solutions have been developed to reduce longitudinal motion distortion. Lee *et al.* [29] suggested scanning the same region with two spatially-separated beams, registering common features between the two images, and estimating longitudinal motion from the registered results. Although effective, this method doubles the acquisition bandwidth needed. Another approach is increasing the A-scan rate and pullback speed to reduce the impact of physiological motion. Wang *et al.* [30] reported an intravascular OCT technique called "Heartbeat OCT", which operated at 2.88 MHz A-scan rate, 4,000 frames/s, and 100 mm/s pullback speed. An entire vessel segment could be scanned within 1 cardiac cycle, reducing the effects of cardiac motion.

However, the circumference of intravascular OCT catheters is much smaller than catheters used in other organs, such as the gastrointestinal tract. Implementing an analogous approach to “Heartbeat OCT” in larger luminal organs would require excessive increases in A-scan rate, which is technically challenging [31,32].

In this paper, we introduce a novel method that uses only a single en face OCT image to estimate and correct for circumferential and longitudinal motion distortion in catheter-based OCT volumes. Circumferential distortion estimation/compensation uses NURD-corrected data [22] to generate an en face OCT image and estimate the circumferential motion distortion. Using NURD-corrected data allows us to minimize the distortion from non-uniform micro-motor rotation, leaving residual distortion from the relative motion between the catheter and tissue. We developed a modified optical flow algorithm to estimate the catheter/tissue displacement from a selected en face image. In contrast to the traditional optical flow [33] method that uses a sequence of images for registration and estimation of the displacement vector, our method only requires a single en face image to estimate the circumferential displacement. It jointly estimates both the displacement and partial derivative of image intensity simultaneously. The en face image is then resampled to correct all remaining circumferential motion distortion. This image is subsequently used for longitudinal distortion correction. The longitudinal distortion correction quantifies the standard deviation of en face image features within a fixed size window sliding along the pullback direction to estimate the longitudinal sampling interval. Then, the en face image is resampled to correct the longitudinal motion distortion. The process is repeated to correct all en face images in the OCT volume. We demonstrate our method for distortion correction on synthetic en face images and OCT images of the esophagus from patients undergoing surveillance endoscopy for Barrett’s esophagus.

## 2. Methods

### 2.1. OCT imaging instrument

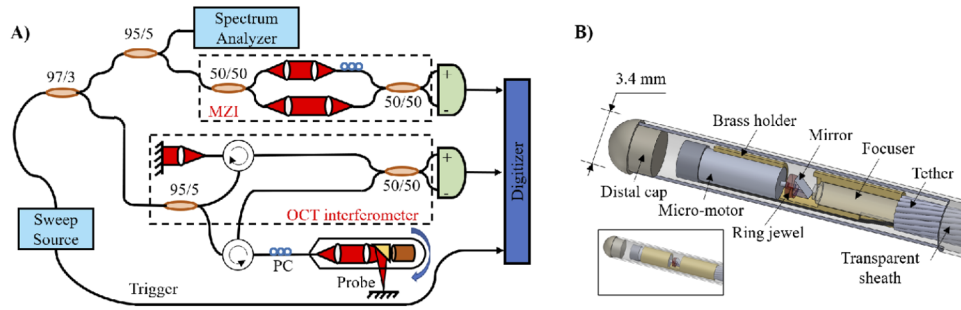
We acquired data using a prototype ultrahigh speed, MEMS-VCSEL SS-OCT system operating at 600 kHz A-scan rate. A detailed description of this system has been previously published [18]. The instrument has 8  $\mu\text{m}$  axial and 20  $\mu\text{m}$  transverse resolution (full width at half maximum), and a 2.4 mm imaging range (in tissue). Volumetric OCT scanning was performed by a micromotor at 24,000 rotations per minute, 400 Hz frame rate, with longitudinal pullback speed of 2 mm/second for 8 seconds, imaging a  $\sim 10\text{ mm} \times 16\text{ mm}$  (circumferential  $\times$  longitudinal) area. A strut on the motor housing blocked  $\sim 40\%$  of the circumferential field of view, leaving an imaging area of  $\sim 6\text{ mm} \times 16\text{ mm}$ . Each B-scan consisted of 1,500 A-scans and each volume had 3,200 B-scans.

Imaging was performed using a side-viewing, micro-motor OCT catheter (Fig. 1(B)). A 2 mm diameter micro-motor (DBL02, Namiki Precision, CA) scanned the OCT beam circumferentially. An optical fiber and graded index lens assembly focused the beam  $\sim 200\text{ }\mu\text{m}$  outside the outer wall of the probe sheath. The fiber / graded index lens was aligned to the scanning micro-motor mirror / prism by a metal housing and attached to a stainless-steel torque coil. The imaging assembly was contained within a high lubricity transparent PTFE sheath (Zeus Inc.) to facilitate smooth longitudinal pullback.

### 2.2. Clinical setting

The study was performed under an IRB protocol approved by VA Boston Healthcare System (VABHS), Harvard Medical School, and Massachusetts Institute of Technology. Written informed consent was obtained before imaging. Data was obtained from patients with Barrett’s esophagus (14 patients with history of dysplasia and 30 with non-dysplasia Barrett’s esophagus (NDBE)), who were undergoing endoscopic surveillance for dysplasia at the VABHS between September 2013 and March 2017. A dual-channel endoscope was used to introduce the OCT probe through





**Fig. 1.** A) Schematic of the ultrahigh speed SS-OCT system. B) Construction of the micro-motor scanning probe with an inset showing an intact view of the probe head.

the endoscope accessory channel and position it on regions of interest. Multiple OCT volumes were acquired from the squamous region, gastroesophageal junction (GEJ), squamocolumnar junction (SCJ), areas with nodularity and/or irregular mucosal/vascular patterns on white-light endoscopy and narrow-band imaging. Additional details about the study setting and patient demographics can be found in Ref. [34].

### 2.3. Overview of the data-processing pipeline

Figure 2 shows a block diagram of the processing pipeline with a breakdown for each block shown in Fig. 3. The pipeline consists of three steps. In Step 1, we process raw OCT fringes to calculate the B-scan OCT intensity image. Next, we correct for NURD using edges of the probe housing as fiducial marks, as described in Ref. [22]. An OCT volume is generated by concatenating the sequential NURD-corrected B-scans. Next, a mean depth projection is performed over  $\pm 50 \mu\text{m}$  depth ( $100 \mu\text{m}$  range) at multiple depths to generate depth-resolved en face images with improved contrast and reduced noise. Finally, the algorithm computes the dynamic ranges of all en face images and chooses one with the largest dynamic range, called image  $I$ . This image is displayed to the user to confirm that it has representative features and there are no data acquisition errors. In Step 2, we estimate the remaining circumferential motion caused by relative motion between the tissue and catheter. We then use the estimated motion to resample  $I$  and generate a second motion-corrected en face image, called  $I_1$ , which should have only longitudinal motion distortion. Finally, in Step 3, we use  $I_1$  to estimate and correct the longitudinal motion. The estimated motion is used to resample all en face images at different depths (without depth projection) to correct the entire volume. We describe the details of Steps 2 & 3 in the next sections.

### 2.4. Estimate and correct circumferential motion distortion

The estimation and correction of circumferential motion distortion is performed iteratively until a predetermined stopping criterion is satisfied (either a maximum number of iterations is reached, or minor improvement is observed with each iteration). Each iteration consists of estimating the circumferential motion and resampling the en face image  $I$  using the estimated motion. We use a maximum of 30 iterations because we found that only minor improvements were obtained with further iterations. We describe details of the sub-steps below and show example correction results using different numbers of iterations.

#### 2.4.1. Estimate circumferential motion from the selected en face image

In the 2D coordinate system of the optimal en face image,  $I$ , let  $c[m, n]$  and  $l[m, n]$  be the circumferential and longitudinal sampling coordinates of a pixel with a circumferential index of  $m$  ( $0 \leq m \leq C - 1$ ) at longitudinal index  $n$  ( $0 \leq n \leq L - 1$ ), where  $C$  and  $L$  are the number

of A-lines in the circumferential dimension and the number of B-frames in the longitudinal dimension, respectively. Because the sampling frequency in the circumferential dimension was much faster (600,000 Hz) than that in longitudinal dimension (400 Hz), we ignore the dependence of  $l$  on the circumferential index  $m$  and refer to it as  $l[n]$ . Hence,  $I\{c[m, n], l[n]\}$  is the image intensity at the pixel  $[m, n]$ . If there is no circumferential motion distortion, all circumferential coordinates with the same index  $m$  will be similar, i.e.  $c[m, n] = c[m, n + k], \forall k = 1, 2, \dots$ . However, because of circumferential motion, this condition is usually not satisfied. Let  $\Delta c[m, n]$  be the circumferential coordinate difference of the pixel index  $[m, n]$  between longitudinal index  $n$  and  $n + 1$ , namely  $\Delta c[m, n] = c[m, n + 1] - c[m, n]$ . We used a modified optical flow algorithm to estimate  $\Delta c[m, n]$  from  $I$ . Using the Taylor's expansion, we can write

$$I\{c[m, n + 1], l[n + 1]\} \approx I\{c[m, n], l[n]\} + \frac{\partial I}{\partial c} \Delta c[m, n] + \frac{\partial I}{\partial l} \Delta l[n] + H.O.T, \quad (1)$$

where H.O.T denotes higher order terms and  $\partial I / \partial c, \partial I / \partial l$  denote partial derivatives of the image intensity in the circumferential and longitudinal directions, respectively. In Eq. (1), both intensities  $I\{c[m, n + 1], l[n + 1]\}$  and  $I\{c[m, n], l[n]\}$  are available. To solve for  $\Delta c[m, n]$ , it is necessary to estimate the partial derivatives  $\partial I / \partial c$  and  $\partial I / \partial l$ . The partial derivative  $\partial I / \partial c$  can be well approximated from the difference in image intensity

$$\frac{\partial I}{\partial c} \approx I\{c[m + 1, n], l[n]\} - I\{c[m, n], l[n]\}. \quad (2)$$

However, the other derivative

$$\frac{\partial I}{\partial l} \approx I\{c[m, n], l[n + 1]\} - I\{c[m, n], l[n]\} \quad (3)$$

cannot be evaluated because the intensity  $I\{c[m, n], l[n + 1]\}$  is not available from the measurement. However, if the displacements  $\Delta c[m, n]$  are known, we can resample the intensity  $I\{c[m, n + 1], l[n + 1]\}$  at  $n + 1$  to obtain  $I\{c[m, n], l[n + 1]\}$ . To resolve the coupling between the two unknowns  $\Delta c[m, n]$  and  $I\{c[m, n], l[n + 1]\}$ , we used a joint estimation algorithm, where one unknown is fixed and the other was updated. For example, when  $\Delta c[m, n]$  is fixed, we can use it to resample  $I\{c[m, n + 1], l[n + 1]\}$  to obtain  $I\{c[m, n], l[n + 1]\}$ , then substitute it into Eq. (3) to compute the partial derivative  $\partial I / \partial l$ . The derivative is used in Eq. (1) to solve for  $\Delta c[m, n]$  and  $\Delta l[n]$  using the Lucas-Kanade (LK) method [33] for optical flow. The solution is used to update  $\Delta c[m, n]$ . This process is repeated until convergence. The method is summarized below (Algorithm 1).

In our implementation, convolutions with Sobel kernels [35] was used to calculate the derivatives  $\partial I / \partial c$  and  $\partial I / \partial l$ . The Sobel kernel effectively computes the differential in one dimension while performing a weighted average in another dimension normal to it, allowing a more robust estimation. Additionally, because the circumferential motion of neighboring pixels is likely very similar, we estimate  $\Delta c[m, n]$  in a coarse grid where only 20 values of the circumferential coordinate  $m$  are used to sample the entire circumference (18 degrees angular sampling pitch). This achieves a significant improvement in algorithm throughput. Next, in order to calculate the circumferential coordinates  $c[m, n]$  from  $\Delta c[m, n]$ , one could simply integrate along the time dimension. However, we found that this integration is sensitive to the estimation error in  $\Delta c[m, n]$ , which sometimes led to unreasonable results for large  $n$ . To avoid this, we enforce additional global constraints on the circumferential coordinates: 1. Top left circumferential coordinate constraint  $c[0, 0] = 0$ . 2. Within one cross-section, the coordinate difference between

---

*Algorithm 1: Estimation of  $\Delta c$  using a modified optical flow method*

Inputs: pixel intensity image  $I$

Outputs: circumferential coordinate change  $\Delta c$

Steps: Let  $I^k\{c[m, n], l[n + 1]\}$  and  $\Delta c^k[m, n]$  be the estimations at the iteration  $k^{th}$ .

1. Initialize  $\Delta c^0[m, n] = 0$  and  $I^0\{c[m, n], l[n + 1]\} \equiv I^0\{c[m, n], l[n]\}$   
(assuming no circumferential distortion).
2. For  $k = 1, 2, \dots$  repeat until convergence:
  - a. Solve for  $\Delta c^{k+1}[m, n]$  using the LK method with image intensities  $I\{c[m, n + 1], l[n + 1]\}$ ,  $I\{c[m, n], l[n]\}$ , and  $I^k\{c[m, n], l[n + 1]\}$ .
  - b. Use  $\Delta c^{k+1}[m, n]$  to resample  $I\{c[m, n + 1], l[n + 1]\}$  to obtain  $I^{k+1}\{c[m, n], l[n + 1]\}$ .
  - c. Check the stopping condition by computing the relative update ratio of  $\delta_{st} = \max_{[m, n]} |\{\Delta c^{k+1}[m, n] - \Delta c^k[m, n]\} / \Delta c^k[m, n]|$ . If  $\delta_{st} < \varepsilon (= 10^{-3})$

or

$k \geq 30$ , exit or else continue.

---

neighboring indices  $m$  should be within a feasible range i.e.

$$[1 - \delta]d \leq c[m + 1, n] - c[m, n] \leq [1 + \delta]d, \quad (4)$$

where  $d$  is the average circumferential distance between two neighboring A-scans.  $\delta$  defines the tightness of the constraint (smaller is more constrained). To enforce a smooth change in  $c$ , we chose a small value,  $\delta = 0.01$ .

These constraints were cast into a convex quadratic programming problem with constraints to solve for  $c[m, n]$  from the estimated  $\Delta c[m, n]$ , namely

$$\mathbf{c}^* = \arg \min_{\mathbf{c}} \sum_{[m, n]} |c[m, n + 1] - c[m, n] - \Delta c[m, n]|^2, \quad (5)$$

subject to  $c[0, 0] = 0$  and  $[1 - \delta]d \leq c[m + 1, n] - c[m, n] \leq [1 + \delta]d$ . Here, we use the bold font to denote vectors (e.g.  $\mathbf{c}$ ) while lower case, italicized font is used for a coordinate of the vector (e.g.  $c[m, n]$ ). The optimization problems were solved using the Optimization Toolbox of MATLAB, Mathworks Inc. As mentioned above, we use  $\delta = 0.01$  in Eq. (4), which is a relatively small number, and perform the correction over multiple iterations that alternate between estimating the circumferential motion and resampling the en face image. Our experimental results show that it is possible to relax this constraint with a larger value for  $\delta$  and reduce the number of iterations. However, we found the convergence was smoother if we clipped the estimated displacements to underestimate the actual displacements and performed “partial” correction over multiple iterations (most converged in 5 to 10 iterations), rather than performing a “full” correction in only 1 or 2 iterations. This is because the Taylor expansion in Eq. (1) is only valid for small displacements  $\Delta c$ . Therefore, it is better to distribute the correction for large  $\Delta c$  over multiple iterations and perform a small correction on each iteration.

#### 2.4.2. Resample the en face image to correct circumferential distortion

After solving for the circumferential coordinates  $\mathbf{c}^*$  of all pixels, we compute the maximum value of the circumferential coordinate across all estimated coordinates  $\mathbf{c}^*$  as  $c_{\max} = \lfloor \max \mathbf{c}^* \rfloor$ . From this, we compute new a new vector of circumferential coordinates  $\mathbf{c}_1$  that uniformly sample the range  $[0, c_{\max}]$  with unit increments, namely  $\mathbf{c}_1 = [0, 1, \dots, c_{\max}]^T$  and  $c_1[m_1] = m_1$  for  $0 \leq m_1 \leq c_{\max}$ . Finally, we use bilinear interpolation to resample the intensity  $I\{c[m, n], l[n]\}$  for each longitudinal coordinate  $n$  to obtain  $I_1\{c_1[m_1], l[n]\}$ . Note that the longitudinal index  $n$  is not required to describe the resampled circumferential coordinate vector  $\mathbf{c}_1$  because it is the same for all  $n$ .

#### 2.4.3. Example results of circumferential distortion correction

Figure 4(A) shows an example of an en face OCT image  $I$  acquired from an NDBE patient which has noticeable distortion of the mucosal surface pattern caused by circumferential motion (mucosal surface patterns in NDBE are typically homogeneous and more circular). Figures 4(B)–4(D) show how the en face image evolves with different number of iterations. Only a fraction of the circumferential motion distortion is corrected (Fig. 4(B)) after 5 iterations, while most of the distortion is corrected after 30 iterations. The insets show 2x zoomed images of a region from the en face images with almost no remaining circumferential distortion after 30 iterations. The motion-corrected image after 30 iterations  $I_1$  was passed to the next step in the algorithm, to estimate and correct longitudinal motion distortion.

### 2.5. Estimate and correct longitudinal motion distortion

The estimation and correction of longitudinal distortion was performed iteratively. In each iteration, an estimation of the longitudinal sampling interval were performed, followed by a resampling step to correct for the longitudinal distortion. A checking step is performed after each iteration to determine if convergence has been reached and the iteration can be terminated.

#### 2.5.1. Estimation of the longitudinal sampling interval

The longitudinal motion distortion is estimated from  $I_1\{c_1[m_1], l[n]\}$ . This type of distortion is caused by variations in the longitudinal sampling interval  $\Delta l[n] = l[n+1] - l[n]$  with  $n$ . To quantify  $\Delta l$ , we begin with the observation that there is a strong correlation between the interval  $\Delta l$  and en face image feature distortion. A shorter value of  $\Delta l$  can be caused by the optical assembly not pulling back with uniform speed, or sticking in the sheath, as well as by the probe sliding in the direction opposite the pullback. This causes the en face image features to appear stretched, because the same features are sampled for a longer time that they should be. Conversely, a longer value of  $\Delta l$  can be caused by the optical assembly rapidly pulling back within the sheath, as well as the probe sliding in the direction of the pullback. This distortion causes the tissue to be more sparsely sampled, resulting in longitudinal compression of features in the en face image. Based on this observation, we estimated the sampling interval by measuring the diversity of the image features over a fixed-sized window sliding along the longitudinal dimension. The diversity is quantified by first performing a one-dimensional, short-time Fourier transform along the longitudinal dimension, with a window of 5 pixels around every pixel of interest ( $\sim 50 \mu\text{m}$  long). The window size was chosen as a compromise between localization and feature diversity for longitudinal motion estimation. A small window provides better localization. However, the feature diversity may be insufficient to ensure reliable estimation. A large window will capture sufficient feature diversity at the cost of less localization. Next, we compute the power spectral density by squaring the magnitude spectrum. Then, we compute a deviation  $\sigma[m_1, n]$  from the power spectral density to characterize feature diversity at each pixel  $[m_1, n]$ . A smaller value of  $\sigma$  means a narrower power spectra or less variation in the image features. Regions where the probe is out-of-contact with the tissue do not contain any image features, therefore the corresponding  $\sigma$

values do not necessarily represent the longitudinal velocity and pixels from these regions are excluded from the longitudinal motion estimation. The out-of-contact regions are identified by a simple thresholding step, where every pixel with intensity less than a threshold is classified as out-of-contact. To estimate  $\Delta l[n]$ , we used the median value of the deviation  $\sigma_{med}[n]$  across the circumferential dimension after excluding all the  $\sigma$  values from out-of-contact regions, namely

$$\sigma_{med}[n] = \text{median}_{m_1} \sigma[m_1, n]. \quad (6)$$

The choice of median deviation provides a good measure of image feature variation across the circumferential dimension as well as robustness to estimation errors of  $\sigma$  from imaging artifacts (e.g. variations in intensity from polarization changes, cross-sections with more than one type of tissue, etc.). More importantly, this method minimizes the estimation error due to naturally elongated tissue structures. These structures are expected to be local instead of global in the circumferential direction. To estimate  $\Delta l[n]$ , we hypothesize that there exists a increasing function  $f$ , such that  $\Delta l[n] = f\{\sigma_{med}[n]\}$ . When  $\Delta l$  is small, there is less variation in the image features, hence,  $\sigma$  is small and vice versa. Furthermore,  $f(0) = 0$  because when the pullback velocity is zero, i.e.  $\Delta l = 0$ , the image feature remains unchanged. Differences in tissue types had a strong influence on the median deviation  $\sigma_{med}$ . For example, squamous mucosa (which is devoid of mucosal surface patterns) was typically much smoother than gastric mucosa or regions of Barrett's esophagus (which have mucosal surface patterns). Hence, squamous mucosa often had smaller  $\sigma_{med}$  than other tissues. To account for the difference in the tissue type, we chose the function  $f(\cdot)$  to be of the following form

$$\Delta l[n] = f\{\sigma_{med}[n]\} = \frac{\sigma_{med}[n]}{\langle \sigma_{med}[n] \rangle_n}, \quad (7)$$

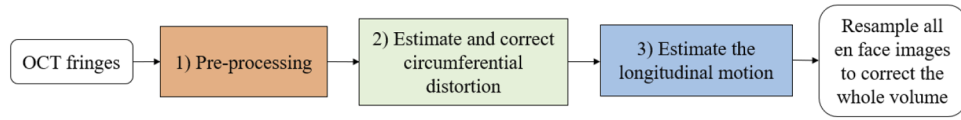
where  $\langle \sigma_{med}[n] \rangle_n$  denotes averaging of  $\sigma_{med}[n]$  over a window larger than the scale of the longitudinal distortion. In our implementation, the window size was chosen to be 100 samples. From the definition in Eq. (7), it can be seen that when  $\sigma_{med}[n] = \langle \sigma_{med}[n] \rangle_n$ , the sampling interval  $\Delta l[n]$  is one pixel and no correction is needed. To estimate the sampling interval, we use the image  $I_1$  to compute the median deviation  $\sigma_{med}[n]$  at each time  $n$ . A running average over a length of 100 samples is used to compute  $\langle \sigma_{med}[n] \rangle_n$ . Then, we substitute both quantities into Eq. (7) to compute the sampling interval  $\Delta l[n]$ . Finally, we clip  $\Delta l[n]$  to be within  $[0.8, 1.2]$ . This prevents excess correction in each iteration and facilitates convergence to a smoother longitudinal distortion correction.

### 2.5.2. Correct longitudinal motion distortion

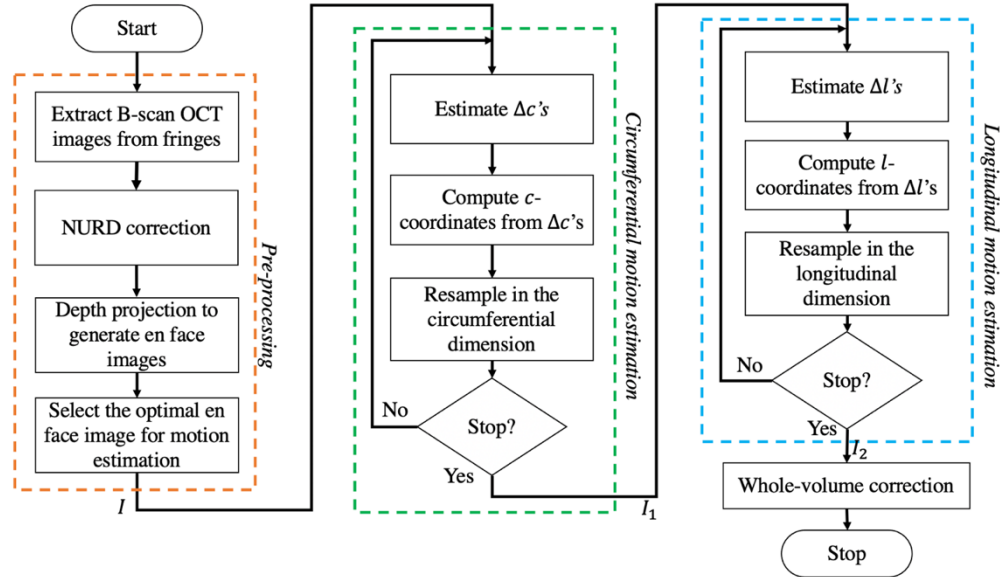
To correct for the longitudinal distortion, we compute the longitudinal sampling coordinates  $l[n]$  as  $l[n] = \sum_{k=0}^{L-1} \Delta l[k]$  and  $l[0] = 0$ . Next, we extract the maximum longitudinal sampling coordinate  $l_{\max}$  as  $l_{\max} = \max_n l[n] = l[L-1]$ . We use this maximum value to compute a set of new longitudinal coordinates  $l_1[n_1]$  that uniformly sample the range  $[0, l_{\max}]$  as

$$l_1[n_1] = \frac{l_{\max} n_1}{L-1}, \quad (8)$$

with  $0 \leq n_1 \leq L-1$ . Finally, we use spline interpolation to resample  $I_1\{c_1[m_1], l[n]\}$  at  $l_1[n_1]$  for each circumferential coordinate  $m_1$  to generate a new image  $I_2\{c_1[m_1], l_1[n_1]\}$  with corrected longitudinal motion distortion. We use spline instead of linear interpolation for longitudinal resampling because the longitudinal distortion is more pronounced than circumferential distortion. Hence, a better interpolator (e.g. spline) is required. A summary of parameters is provided in Table 1.



**Fig. 2.** Processing pipeline to correct for circumferential and longitudinal motion distortion.



**Fig. 3.** Breakdown of different blocks in the processing pipeline.

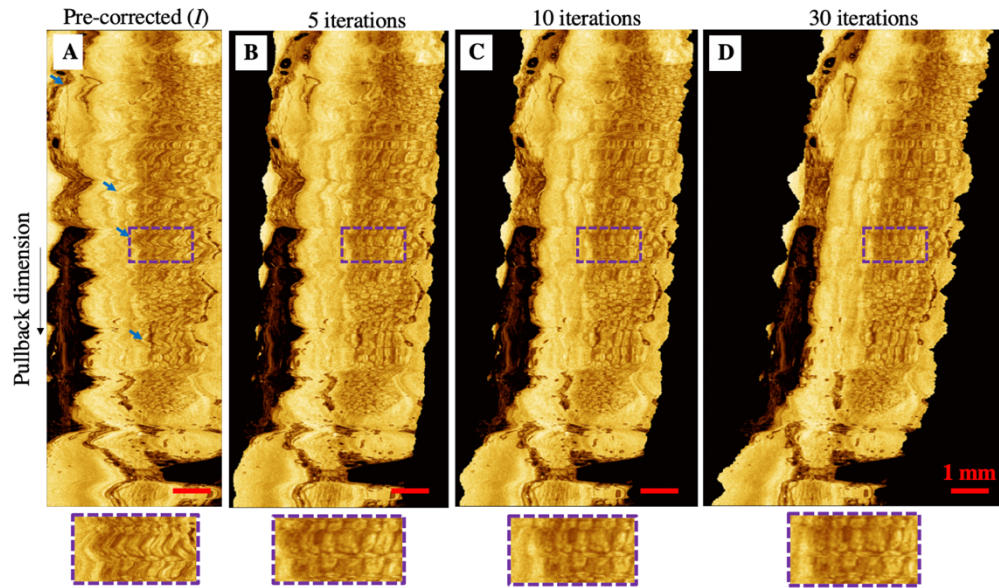
**Table 1. Parameters used in motion correction algorithm**

Parameters	Unit	Values	Effect on performance
Maximum number of iterations for circumferential distortion estimation and correction	Iterations	30	Convergence speed
Circumferential coordinate tightness constraint factor ( $\delta$ )	Pixels	0.01	Correction quality
Window length to evaluate feature diversity in longitudinal sampling interval estimation ( $L_{features}$ )	Pixels	5	Correction quality
Window length [pixels] to normalize the difference in tissue type ( $L_{tissue-normalizing}$ )	Pixels	100	Correction quality
Allowable range of $\Delta l$ in each iteration	Pixels	[0.8, 1.2]	Convergence speed
Difference between 10% and 90% percentiles of $\Delta l$ to stop longitudinal sampling interval estimation	Pixels	0.03	Convergence speed
Maximum number of iterations for longitudinal sample interval estimation	Iterations	300	Convergence speed

### 2.5.3. Stopping criteria for the longitudinal correction

After each iteration, a test is performed to determine a stopping condition. The sampling values  $\Delta l[0], \dots, \Delta l[L-1]$  are sorted from minimum to maximum. Then, the difference between 10% and 90% percentiles is computed and compared to a predetermined threshold (selected to be 0.03). If the difference is less than the threshold, the iteration is terminated. Otherwise, we





**Fig. 4.** Correction results for circumferential motion distortion versus different number of iterations. The blue arrows point to areas with pronounced distortion due to circumferential motion.

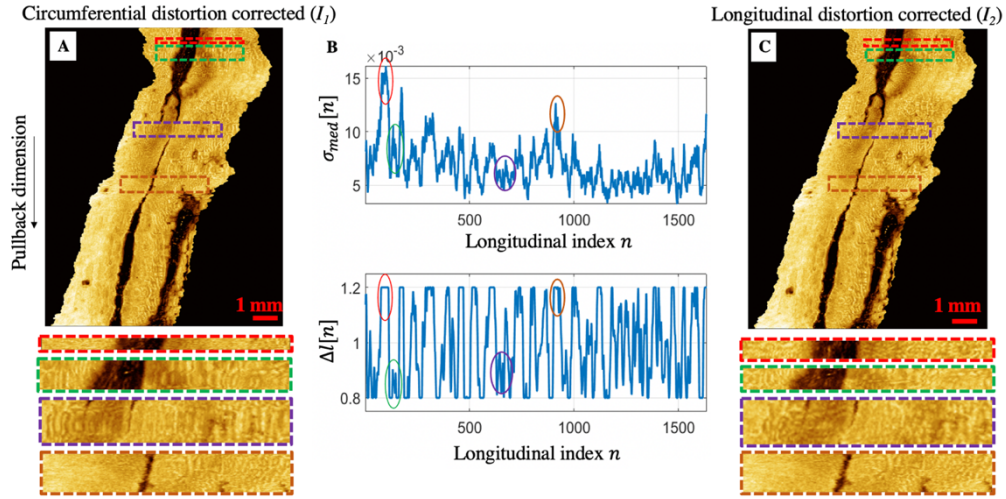
continue iteration until the stopping condition is satisfied or the maximum number of iterations (set to 300) is reached. Convergence is usually achieved within the first 100 iterations. Our source code for image distortion correction is at: [https://gitlab.com/huutan86/distortion\\_correction.git](https://gitlab.com/huutan86/distortion_correction.git).

#### 2.5.4. An example of longitudinal distortion correction

Figure 5(A) shows an example of an en face image  $I_1$  of the esophagus from a NDBE patient. This image was corrected for circumferential distortion using the procedure in Section 2.4. Using this image, we calculated the median deviation of the image features as shown in Fig. 5(B) (top). This deviation was used to estimate the sampling interval  $\Delta l[n]$  from Eq. (7). The estimated sampling interval is shown in Fig. 5(B) (bottom). Different regions with obvious longitudinal distortion in Fig. 5(A) are marked by colored rectangles with an zoomed insets in the bottom. There were regions with both compression (red and brown) and stretching (green and purple) of en face features. The corresponding values for the median deviation  $\sigma_{med}[n]$  and sampling interval  $\Delta l[n]$  are annotated with matching colors in Fig. 5(B). From this plot, one can see that the compressed regions (red and orange ovals) have larger median deviations and sampling intervals, while the stretched regions (green and purple ovals) have smaller median deviations and sampling intervals. The estimated sampling intervals were used for resampling the en face image to compensate for the effects of longitudinal distortion. This process was repeated until convergence (18 iterations were needed in this example). The final en face image ( $I_2$ ) is shown in Fig. 5(C) with most of the longitudinal distortion removed. The 3x zoomed insets at the bottom of Fig. 5(A) and Fig. 5(C) show that stretched features are restored, demonstrating that longitudinal motion distortion is corrected.

#### 2.6. Motion distortion correction of the entire OCT volume

To correct the entire OCT volume, we first sliced it into multiple en face images (without depth projection). For each en face image, we correct the circumferential distortion, followed by



**Fig. 5.** A) A circumferential motion distortion corrected image ( $I_1$ ) used for longitudinal motion estimation. The insets in the bottom shows a 3x zoomed regions of the en face image. B) Median deviation of the image features  $\sigma_{med}[n]$  (top) and estimated  $\Delta I[n]$  (bottom) as the functions of longitudinal index  $n$  in the 1<sup>st</sup> iteration. The ovals in these plots show the estimated  $\sigma_{med}[n]$  and  $\Delta I[n]$  for different regions in the en face image in A) (color matched). C) En face image corrected for longitudinal distortion; the insets show 3x zoomed regions. A clear reduction of the longitudinal distortion can be observed.

longitudinal distortion. The correction for circumferential motion distortion was performed iteratively, following the procedure in Section 2.4. However, here we use the estimated displacements from Section 2.4.1 for resampling and do not re-estimate the circumferential motion for each en face image. The correction for longitudinal motion distortion is similar to that described in Section 2.5, except that we use the estimated sampling interval in Section 2.5.1 for resampling and do not re-estimate the sampling interval.

### 3. Results

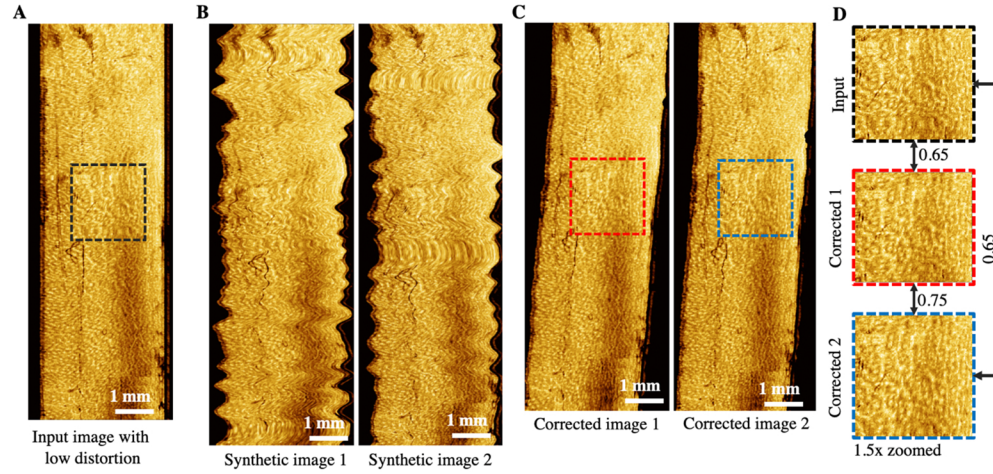
#### 3.1. Demonstration of distortion correction on synthesized test images

It is difficult to evaluate the algorithm performance for endoscopic OCT because ground truth data is not available. Therefore, we tested performance by using an input volume with minimal motion distortion (Fig. 6(A)) and synthesized two motion distorted test volumes to mimic distortion commonly observed in clinical data. The circumferential distortion was generated using the formula

$$\Delta c[m, n] = g[(h * x)][n], \quad (9)$$

where  $x[n]$  was drawn from a normal distribution for each longitudinal index  $n$ ,  $h[n] = (1/M)[e^{-0.5(n/15)^2}]$  was a normalized Gaussian filtering kernel to simulate the degree of the correlation between longitudinal indices under circumferential distortion.  $M$  was a normalizing constant, selected so that  $\sum_n h[n] = 1$ . The standard deviation of this Gaussian was selected to be 15 pixels, chosen based on distortion seen in typical clinical data. Finally, the normalizing function  $g(\cdot)$  scaled the filtered signal amplitude so that the maximum amplitude of  $\Delta c$  was 50 pixels, a value selected based on the clinical data. Using the generated  $\Delta c[n]$ , we shifted and

zero-padded each longitudinal index  $n$  to generate a circumferentially distorted image. This image was then processed to simulate longitudinal motion distortion.



**Fig. 6.** A) The input image with low motion distortion used to synthesize distorted test images, B) Synthetic test images generated from A) using two random motion profiles, C) Distortion corrected test images from B). The boxes denote regions used to compute the structure similarity index measures (SSIM) between the low distortion input image, and distortion corrected test images. D) 1.5x zoomed regions of the patches used for calculating the SSIM. The arrows between the patches denote the SSIM values.

To simulate the longitudinal motion distortion, we used the equation

$$\Delta l[n] = 1.0 - \sum_{i=1}^5 0.8e^{-0.5[(n-n_i)/30]^2}, \quad (10)$$

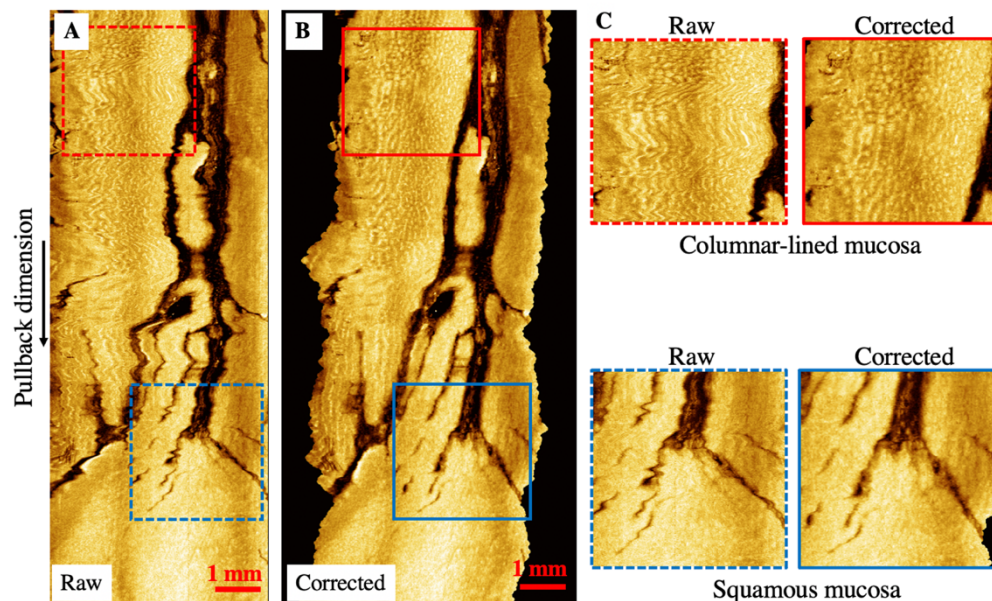
where  $n_i (i = 1, \dots, 5)$  were five random positions along the pullback. These locations were randomly sampled, such that  $\Delta l$  was positive in both cases. This mimics a situation where the longitudinal pullback of the optical assembly had sticking 5 times. When the pullback speed was reduced by sticking, the longitudinal sampling slows down from the maximum value of 1.0 to 0.2 (= 1.0 – 0.8). Using this  $\Delta l[n]$ , we resampled the circumferentially distorted image generated in the previous step to add longitudinal distortion. The distorted test images are shown in Fig. 6(B). Then, the motion correction was applied to these test images to remove the distortion. Corrected images are shown in Fig. 6(C). Note that the two motion corrected images are very similar to the input image, even though the distorted test images are synthesized using very different motion profiles. This demonstrates the convergence and stability of the algorithm.

To quantify the similarity between correction outputs, we extracted corresponding test patches from the input image and the two motion corrected test images as shown in Fig. 6(D). The patches were compared using structure similarity index measures (SSIMs) [36], which is a number from 0 to 1 telling the similarity between the patches using the luminance, contrast, and structures. A higher value of SSIM corresponds to greater similarity. Note that the SSIM between the two motion corrected patches is high (0.75), suggesting the convergence of the motion corrected output to the same result, irrespective of the motion profile that was applied to the input image. This value is larger than the SSIM between each motion corrected patch and the input image, suggesting that there may have been a small amount of motion in the input image that was corrected by the algorithm.



### 3.2. Motion distortion correction of OCT images of human esophagus

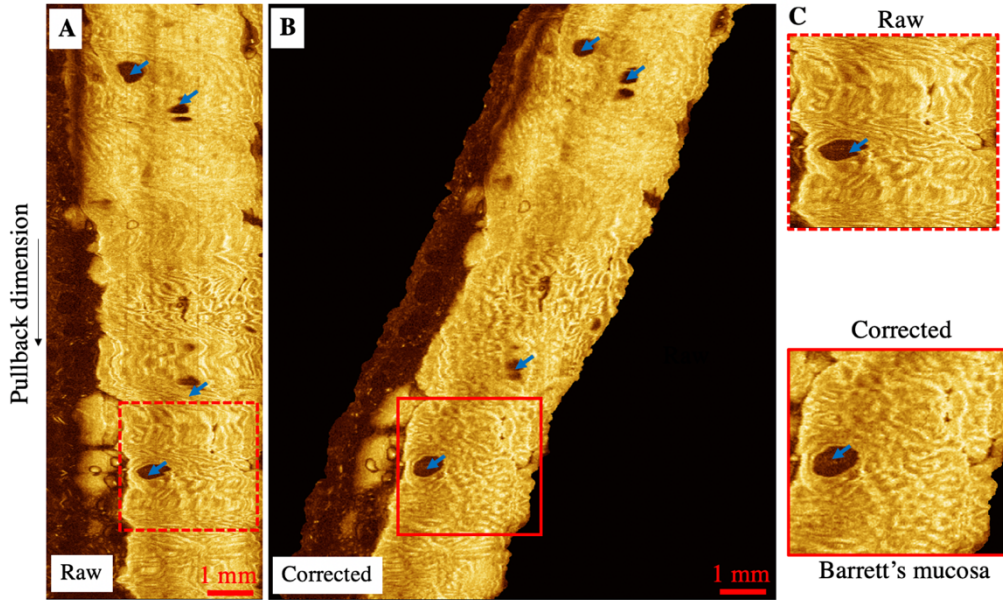
Next, we validated the motion distortion correction algorithm on a selection of en face images from patients with different pathologies. Figure 7(A) shows an en face OCT image (at  $\sim 60\ \mu\text{m}$  below the mucosal surface) from a NDBE patient under surveillance for dysplasia with in a region affected by motion distortion. The motion-corrected image is shown in Fig. 7(B). Figure 7(C) shows selected 1.5x zoomed regions of the NDBE mucosa (red) and squamous mucosa (blue). The motion distortion of the columnar epithelium mucosal patterns (top-left inset) was corrected (top-right inset). The bottom left image has a region of homogeneous squamous epithelium with darker areas caused by insufficient or out-of-contact tissue. There is a high degree of waviness in the boundaries of these areas (bottom-left inset) caused by rapid circumferential motion distortion. This distortion was almost completely corrected (bottom-right inset), facilitating a more accurate evaluation of mucosal features.



**Fig. 7.** A) En face OCT image from a patient with NDBE, B) Distortion corrected version of A). C) 1.5x zoomed images of regions in A) and B).

We then assessed the algorithm performance on images from patients with dysplastic BE. Figure 8(A) shows an en face image (at  $\sim 40\ \mu\text{m}$  below the mucosal surface) with low-grade dysplasia. Blue arrows indicate hypo-scattering atypical glands, characteristic OCT features of dysplasia [34]. The mucosal surface pattern in the dashed red region shows distorted mucosal surface patterns associated with dysplastic BE. Both circumferential and longitudinal motion distortion is visible in this region. Longitudinal motion distortion causes the mucosal features to appear stretched ( $\sim 1/3$  from the top) and compressed (about half of the region, right of hypo-scattering structure). Figure 8(B) shows the motion corrected image.

Figure 8(C) shows regions of interest enlarged by 1.5x. After motion correction, the mucosal surface pattern is more homogeneous with no noticeable compression or stretching. Also, glandular structures appear less distorted after motion correction. Before motion correction, the gland boundary has a discontinuity with sharp edges, which was likely caused by rapid circumferential motion distortion and does not represent the true shape. Distortion and irregularity



**Fig. 8.** En face image from a patient with low grade dysplasia, B) Distortion corrected version of A). C) 1.5x zoomed images of a region with irregular mucosal surface pattern in A) and B).

in mucosal surface patterns and glandular structures are possible OCT markers of dysplasia, therefore motion-corrected images can reduce the risk of interpretation errors.

Figure 9 shows several correction examples different regions extracted from the clinical images. These regions represent different features observed in our dataset including: small pit pattern of columnar epithelium (Fig. 9(A)), small Barrett's esophagus (BE) island (Fig. 9(B)), large pit pattern with large (Fig. 9(C)) and small (Fig. 9(D)) amounts of motion, squamous epithelium (Fig. 9(E)), and the SCJ function (Fig. 9(F)). Consistent reduction in the motion distortion is observed for all cases.

### 3.3. Quantification of the performance for distortion correction using clinical data

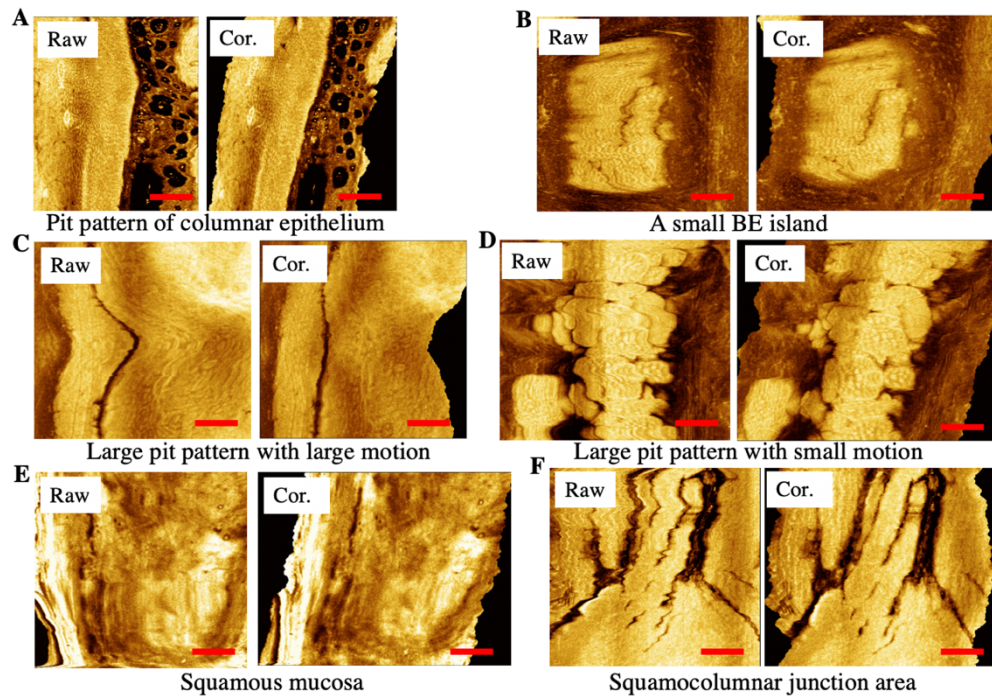
To benchmark the performance of the algorithm across all datasets, we use two different metrics to evaluate the amount of residual circumferential and longitudinal motion distortion, respectively. The first metric is the average magnitude of the circumferential median  $med_m \Delta c[m, n]$  along the longitudinal dimension.

$$d_{\Delta c} = \langle |med_m \Delta c[m, n]| \rangle_n. \quad (11)$$

The circumferential median  $med_m \Delta c[m, n]$  was used to quantify the amount of circumferential distortion for each longitudinal index  $n$ . The magnitude was used because this quantity can be negative or positive. A small value of  $d_{\Delta c}$  corresponds to a lower amount of circumferential distortion and vice versa. The second metric is the standard deviation of  $\Delta l[n]$  along the longitudinal dimension.

$$d_{\Delta l} = \sigma_n \{ \Delta l[n] \}. \quad (12)$$

A small value of  $d_{\Delta l}$  corresponds to a more uniform sampling in the longitudinal dimension i.e. lesser amount of residual longitudinal motion and vice versa. Figure 10 shows two scatter plots for  $d_{\Delta c}$  and  $d_{\Delta l}$  from 54 datasets obtained from all 44 patients. One data point corresponds to 1 dataset with the color indicating the diagnosis. A dashed black line corresponding to the



**Fig. 9.** Representative regions of the esophagus before (left) and after (right) correction. Scale bar: 1 mm.

equation  $y = x$  is drawn in each plot dividing the plot into two domains where the method was effective (below the line) vs. not effective (above the line).

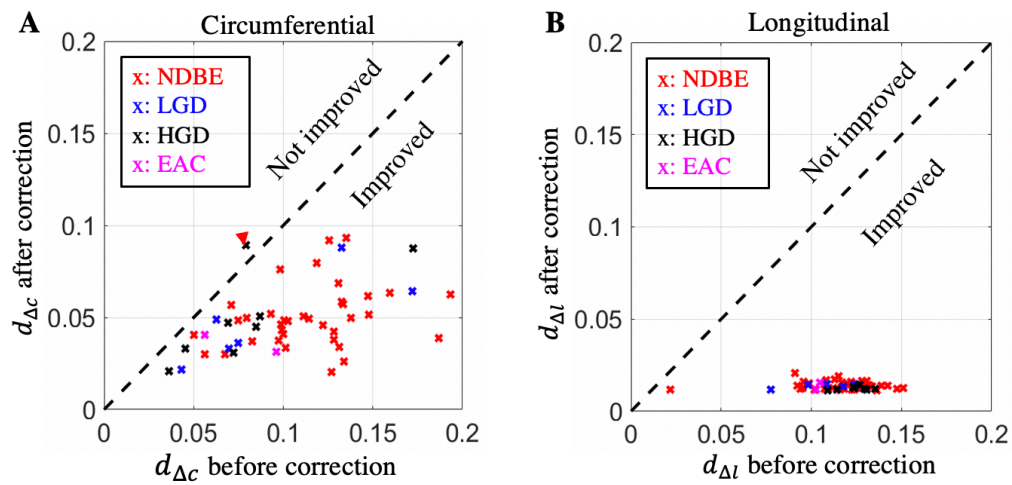
We notice that the amount of circumferential distortion varied appreciably among these datasets, but there were no large differences in motion corrected results between subgroups. While some datasets had a very small amount of circumferential distortion  $d_{\Delta c} \approx 0.05$  pixel, others had a much larger amount of  $d_{\Delta c} \approx 0.2$  pixel. Circumferential motion distortion was reduced in most of the datasets, except for one (arrowhead). Figures 11(A)–11(B) shows the en face images of this dataset before and after motion distortion correction. The 1.5x zoomed regions (Fig. 11(C)) of these en face images show very complicated features with a longitudinal compression coupled with possible circumferential displacement. The algorithm attempted to correct distortion by stretching the image in the longitudinal dimension and correcting for circumferential displacement. The corrected image did not appear to improve, likely due to improper sampling in the raw en face image. In contrast to circumferential motion distortion, a reduction in longitudinal motion distortion was observed in all the datasets, with a saturated performance at  $d_{\Delta l} \leq 0.02$  pixel for all cases.

### 3.4. Processing time

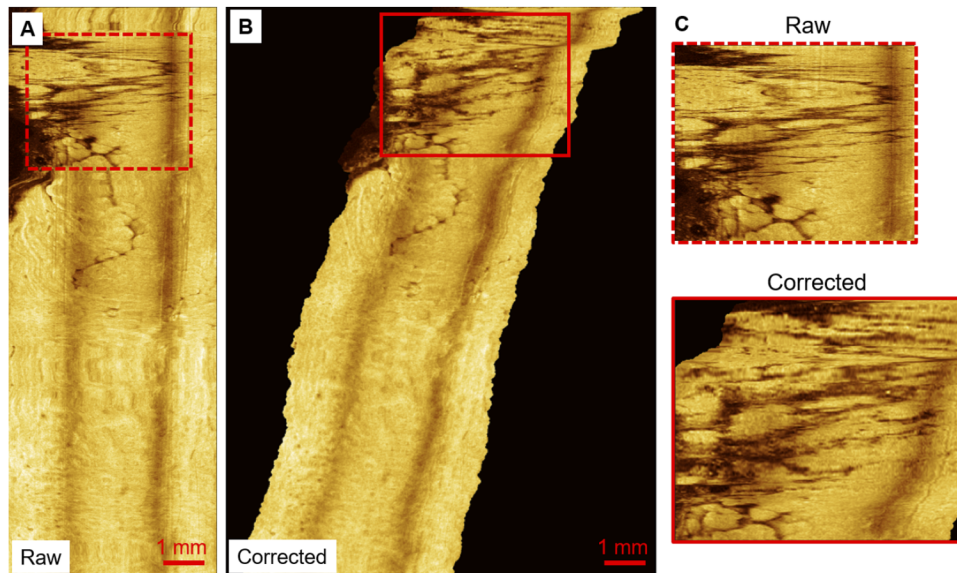
Table 2 shows processing time for different steps of the algorithm to process an OCT volume of 3,200 B-scans with 1,500 A-scans per B-scan and 240 points per A-scan. Pre-processing to time required to process raw OCT fringes to calculate the B-scan OCT and correct for NURD is not included.

The algorithm requires ~13.5 minutes to process a volumetric OCT data ( $1,500 \times 3,200 = 4.8$  million A-scans). In many cases, the total processing time is less than 13.5 minutes because the stopping condition is reached before the maximum number of iterations. A large portion





**Fig. 10.** Algorithm performance for motion correction in the circumferential (left) and longitudinal directions (right) for 54 datasets from 44 patients. Each “x” symbol corresponds to one dataset. The color of the symbol denotes the patient diagnosis. Dashed lines in each plot correspond to  $y = x$  and separate the plot into two regions showing motion distortion correction performance.



**Fig. 11.** A-B) En face images from a HGD patient where our method failed to reduce the motion distortion. C) 1.5x zoomed images of a region specified in A) and B) show improper sampling due to fast jumping in the pullback and a potentially coupled circumferential distortion.

**Table 2. Processing time for different steps in the motion distortion correction algorithm.**

Processing blocks	Block sub-steps	Avg. time per iteration [s]	Typical # of iterations	Max time [s]
Estimate and correct circumferential distortion.	Estimate $\Delta c$ from $I$ .	13.3	30	399.0
	Solve $c$ from $\Delta c$ .	3.1	30	93.0
	Resampling en face to prepare for the next iteration.	0.22	30	6.6
Estimate and correct longitudinal distortion.	Estimate $\Delta l$ from $I_1$ .	0.73	100	73
	Resampling en face to prepare for the next iteration	0.16	100	16
Whole volume correction.	Load NURD corrected data.	N.A.	N.A.	6.2
	Resample enface images in the circumferential dimension.	0.24	240	57.6
	Resample enface images in the longitudinal dimension.	0.17	240	40.8
	Write result to hard drive.	N.A.	N.A.	109
Total time				801.2 (~13.5 minutes)

(~50%) of the total time is required to estimate  $\Delta c$  from the input en face image  $I$ . The code is in prototype form and for easy of development. Therefore, this step is implemented on CPUs. The estimation is sequentially performed from the most distal to most proximal position of the longitudinal pullback. However, since there is no dependency between estimation of  $\Delta c$  from  $I$  at different longitudinal index  $n$ , this estimation step can be performed in parallel processing using Graphic Processing Units (GPUs). Similarly, estimation of the longitudinal sampling interval  $\Delta l[n]$  can be performed in parallel for different longitudinal index  $n$ .

Volumetric correction with different images at different en face depths, currently requires ~12% of the total time and could be performed in parallel processing. The I/O time to write processed images cannot be easily speeded up by parallel processing. However, this step is only required for off-line data investigation and storage.

#### 4. Discussion

Although the algorithm uses several parameters, many are determined by the instrument and scanning protocol, so the number of arbitrary parameters is not excessive. Algorithm performance depends on a small set of parameters ( $\delta, L_{\text{features}}, L_{\text{tissue-normalizing}}$ ) related to the size of the image feature being evaluated. Other parameters such as  $\Delta l$ , the stopping condition, and the maximum number of iterations only affect the convergence speed rather than the final output. Therefore, the algorithm can be applied to other datasets with a different sampling pitch or different features by appropriately scaling these feature size parameters, while retaining the default values for other parameters. The last column of Table 1 summarizes the effects of parameters on algorithm performance.

Our algorithm utilizes a single en face image to estimate motion and differs from other correction methods that only use cross-sectional images [23,25]. There are two types of motion correction methods using cross-sectional images. The first type estimates the motion using fiducials on the catheter [22]. This approach corrects NURD due to scanning non-uniformity but does not account for lateral tissue motion. While the motion corrected data has minimal rotational distortion, it may still suffer from distortion due to tissue motion. The second type of approach estimates and corrects motion using cross-sectional images of tissue instead of fiducials

[23,25]. This approach accounts for tissue motion in the circumferential direction. However, it does not account for longitudinal motion distortion which causes variation in the longitudinal sampling interval. Our method adds an extra layer of correction on top of NURD correction to enforce consistency in both the circumferential and longitudinal dimensions, which is critical for reliable image interpretation. It was also possible to perform motion distortion correction starting from the en face image generated directly from non-NURD-corrected OCT data. However, we found that using the NURD-corrected data was superior because the estimated circumferential displacement was smaller than if the non-NURD-corrected data was used. This facilitates a smoother convergence of the optical flow algorithm, because the Taylor expansion in the optical flow algorithm performs better for smaller displacement values.

One disadvantage of our algorithm is that it relies on en face image features for motion estimation. We selected en face images with the highest dynamic range, but they typically only contain a subset of features from the OCT volume. Also, the depth selection of the en face image may be optimal for some regions and not others. Images from homogenous regions, such as those with squamous epithelium, will have less contrast than other regions (e.g. Barrett's epithelium), making accurate estimation challenging. While we accounted for this phenomenon using the normalization scheme in Eq. (7), we envision that estimation accuracy can be improved if an optimal depth is selected at each longitudinal (pullback) location instead of using a single selection for all locations. Future increases in A-scan rates will increase the sampling density and the visibility of features used for motion estimation, improving algorithm performance.

Improving tissue coverage will likely increase the robustness and the accuracy of the motion estimation. In our study, we used an endoscope-based catheter of 3.4 mm diameter, which was much smaller than the esophageal lumen. The en face image used for motion estimation only covered a portion of the esophagus, while other regions were out-of-contact. Therefore, the estimated motion was only local to the tissue area imaged by the probe. Improving the tissue coverage will allow a larger portion of the esophagus to be imaged, increasing robustness of the motion estimation. We emphasize that our algorithm should be compatible with modalities which have larger tissue coverage such as OCT capsules [37–39] or OCT balloons [40,41]. However, caution is required to ensure that the en face image has sufficient sampling density to visualize features for motion estimation.

It is important to note that the dominant motion distortion that requires correction is in the transverse direction. Axial motion where tissue is in contact with the probe is relatively small. Images from out of contact regions have poor quality and are usually not interpretable. The OCT probe sheath acts analogously to a microscope slide cover glass, reducing tissue surface roughness which produces aberrations, enabling subsurface imaging.

The problem of motion correction is similar to the wobbling correction problem in computer vision caused by rolling shutters [28,42–44], although the image distortion here was more severe. In endoscopic OCT, sequential cross-sectional B-scans were obtained at different times, causing circumferential motion distortion. In the wobbling correction problem, different rows of the camera sensor are exposed at different times, causing skew and curvature distortion. The similarity between these two problems allows adaptation of wobbling correction methods to correct circumferential distortion. Recently, Rengarajan *et al.* [28] suggested using convolutional neural networks to estimate the row-wise camera motion from image features contained in a single rolling shutter image. Although simple and powerful, this method required a large training dataset that contains hundreds of thousands of synthetic images generated from undistorted images. This requirement is challenging in clinical OCT applications such as endoscopy where the amount of data is limited patient enrollment and procedural complexity. Therefore, a method that requires less training data is preferred.

There are also important differences between the catheter-based OCT problem and the wobbling correction problem. First, the B-scan frame rate for catheter-based OCT is relatively slow (a few

hundred Hertz in endoscopic OCT) compared to the row reading rate in commercial cameras (~65 kHz for a 60 fps 1080p camera). Therefore, circumferential motion distortion in catheter-based OCT is more severe than wobbling from the rolling shutter. Furthermore, the catheter-based OCT problem is confounded by the coupling of longitudinal motion distortion and non-uniform longitudinal sampling.

Our motion distortion correction algorithm can be applied to other catheter/endoscope-based OCT methods such as intravascular OCT, unsedated capsule imaging for Barrett's esophagus screening [39,45–47], or balloon-based OCT esophageal imaging for large circumferential coverage [41,48]. It could also be applied to correct for motion artifacts in lower GI imaging for anal and rectal imaging of mucosal crypt architecture [49].

Machine learning algorithms typically rely on large amounts of data with labels obtained from multiple sources to reduce inter-observer variation. Training these algorithms to account for motion distortion can be costly, because it requires a large amount of clinical OCT images. Furthermore different readers can classify the same features as regular or irregular because of varying tolerance to motion distortion and previous training experience. Motion distortion correction can be used to normalize the data, hence, reduce the amount of data needed to train machine learning algorithms. This is because mucosal features will be less distorted after correction compared to original data. For human readers, evaluation of features such as irregular mucosal surface patterns or atypical glands which are potential markers of BE with dysplasia should be less prone to inter-observer variation [34].

## 5. Conclusions

This paper presents a new method to estimate and correct motion distortion in catheter/endoscope-based OCT. The algorithm begins with OCT pre-processing to generate the OCT volume from raw fringes and correct for NURD using fiducial markers on the probe housing. Then, we select an en face image with high dynamic range and representative features to estimate the motion distortion. This en face image is analyzed using a modified optical flow algorithm to estimate the relative circumferential motion between the tissue and catheter, then resampled to correct circumferential motion distortion. The corrected image is then used to estimate the longitudinal motion distortion from the standard deviation of the image features in a sliding window along the longitudinal direction. After estimating both circumferential and longitudinal motion, all en face images of the NURD corrected volume are resampled to correct the entire volume. We demonstrate the performance of the algorithm on synthetic images and clinical OCT images of the human esophagus from patients with Barrett's esophagus undergoing surveillance for dysplasia. The algorithm can be used in multiple OCT applications which may have circumferential and longitudinal motion distortion.

## Funding

National Institutes of Health (R01-CA075289, R01-CA252216, R44CA235904).

## Acknowledgements

We would like to thank Marisa Figueiredo and Dr. Qin Huang from VABHS in recruiting patients and pathology reading. J. Zhang acknowledges support from the National Science Foundation Graduate Research Fellowship program. Dr. K. Liang acknowledges support from the Agency for Science, Technology and Research (Singapore) graduate fellowship.

## Disclosures

The authors declare no conflicts of interest.



## References

1. G. J. Tearney, M. E. Brezinski, J. G. Fujimoto, N. J. Weissman, S. A. Boppart, B. E. Bouma, and J. F. Southern, "Scanning single-mode fiber optic catheter–endoscope for optical coherence tomography: erratum," *Opt. Lett.* **21**(12), 912 (1996).
2. W. Jung, D. T. McCormick, J. Zhang, L. Wang, N. C. Tien, and Z. Chen, "Three-dimensional endoscopic optical coherence tomography by use of a two-axis microelectromechanical scanning mirror," *Appl. Phys. Lett.* **88**(16), 163901 (2006).
3. Y.-H. Seo, K. Hwang, H.-C. Park, and K.-H. Jeong, "Electrothermal MEMS fiber scanner for optical endomicroscopy," *Opt. Express* **24**(4), 3903 (2016).
4. M. J. Cobb, X. Liu, and X. Li, "Continuous focus tracking for real-time optical coherence tomography," *Opt. Lett.* **30**(13), 1680 (2005).
5. L. Huo, J. Xi, Y. Wu, and X. Li, "Forward-viewing resonant fiber-optic scanning endoscope of appropriate scanning speed for 3D OCT imaging," *Opt. Express* **18**(14), 14375 (2010).
6. N. Zhang, T.-H. Tsai, O. O. Ahsen, K. Liang, H.-C. Lee, P. Xue, X. Li, and J. G. Fujimoto, "Compact piezoelectric transducer fiber scanning probe for optical coherence tomography," *Opt. Lett.* **39**(2), 186–188 (2014).
7. K. Liang, O. O. Ahsen, Z. Wang, H.-C. Lee, W. Liang, B. M. Potsaid, T.-H. Tsai, M. G. Giacomelli, V. Jayaraman, and H. Mashimo, "Endoscopic forward-viewing optical coherence tomography and angiography with MHz swept source," *Opt. Lett.* **42**(16), 3193–3196 (2017).
8. L. M. Wurster, R. N. Shah, F. Placzek, S. Kretschmer, M. Niederleithner, L. Ginner, J. Ensher, M. P. Minneman, E. E. Hoover, H. Zappe, W. Drexler, R. A. Leitgeb, and Ç Ataman, "Endoscopic optical coherence tomography angiography using a forward imaging piezo scanner probe," *J. Biophotonics* **12**(4), e201800382 (2019).
9. J. Wu, M. Conry, C. Gu, F. Wang, Z. Yaqoob, and C. Yang, "Paired-angle-rotation scanning optical coherence tomography forward-imaging probe," *Opt. Lett.* **31**(9), 1265–1267 (2006).
10. H.-C. Park, C. Song, M. Kang, Y. Jeong, and K.-H. Jeong, "Forward imaging OCT endoscopic catheter based on MEMS lens scanning," *Opt. Lett.* **37**(13), 2673–2675 (2012).
11. A. M. Rollins, R. Ung-arunyawee, A. Chak, R. C. K. Wong, K. Kobayashi, M. V. Sivak, and J. A. Izatt, "Real-time in vivo imaging of human gastrointestinal ultrastructure by use of endoscopic optical coherence tomography with a novel efficient interferometer design," *Opt. Lett.* **24**(19), 1358 (1999).
12. X. D. Li, S. A. Boppart, J. Van Dam, H. Mashimo, M. Mutinga, W. Drexler, M. Klein, C. Pitris, M. L. Krinsky, M. E. Brezinski, and J. G. Fujimoto, "Optical coherence tomography: Advanced technology for the endoscopic imaging of Barrett's esophagus," *Endoscopy* **32**(12), 921–930 (2000).
13. B. E. Bouma, G. J. Tearney, C. C. Compton, and N. S. Nishioka, "High-resolution imaging of the human esophagus and stomach in vivo using optical coherence tomography," *Gastrointest. Endosc.* **51**(4), 467–474 (2000).
14. Y. Chen, A. D. Aguirre, P.-L. Hsiung, S. Desai, P. R. Herz, M. Pedrosa, Q. Huang, M. Figueiredo, S.-W. Huang, and A. Koski, "Ultrahigh resolution optical coherence tomography of Barrett's esophagus: preliminary descriptive clinical study correlating images with histology," *Endoscopy* **39**(07), 599–605 (2007).
15. S. H. Yun, G. J. Tearney, B. J. Vakoc, M. Shishkov, W. Y. Oh, A. E. Desjardins, M. J. Suter, R. C. Chan, J. A. Evans, I. K. Jang, N. S. Nishioka, J. F. De Boer, and B. E. Bouma, "Comprehensive volumetric optical microscopy in vivo," *Nat. Med.* **12**(12), 1429–1433 (2006).
16. L. P. Hariri, M. B. Applegate, M. Mino-Kenudson, E. J. Mark, B. D. Medoff, A. D. Luster, B. E. Bouma, G. J. Tearney, and M. J. Suter, "Volumetric optical frequency domain imaging of pulmonary pathology with precise correlation to histopathology," *Chest* **143**(1), 64–74 (2013).
17. P. H. Tran, D. S. Mukai, M. Brenner, and Z. Chen, "In vivo endoscopic optical coherence tomography by use of a rotational microelectromechanical system probe," *Opt. Lett.* **29**(11), 1236–1238 (2004).
18. T.-H. Tsai, B. Potsaid, Y. K. Tao, V. Jayaraman, J. Jiang, P. J. S. Heim, M. F. Kraus, C. Zhou, J. Hornegger, H. Mashimo, A. E. Cable, and J. G. Fujimoto, "Ultrahigh speed endoscopic optical coherence tomography using micromotor imaging catheter and VCSEL technology," *Biomed. Opt. Express* **4**(7), 1119 (2013).
19. J. Li, M. de Groot, F. Helderma, J. Mo, J. M. A. Daniels, K. Grünberg, T. G. Sutedja, and J. F. de Boer, "High speed miniature motorized endoscopic probe for optical frequency domain imaging," *Opt. Express* **20**(22), 24132 (2012).
20. T. Wang, W. Wieser, G. Springeling, R. Beurskens, C. T. Lancee, T. Pfeiffer, A. F. W. Van der Steen, R. Huber, and G. van Soest, "Ultrahigh-speed intravascular optical coherence tomography imaging at 3200 frames per second," *Opt. InfoBase Conf. Pap.* **38**(10), 1715–1717 (2013).
21. W. Kang, H. Wang, Z. Wang, M. W. Jenkins, G. A. Isenberg, A. Chak, and A. M. Rollins, "Motion artifacts associated with in vivo endoscopic OCT images of the esophagus," *Opt. Express* **19**(21), 20722 (2011).
22. O. O. Ahsen, H.-C. Lee, M. G. Giacomelli, Z. Wang, K. Liang, T.-H. Tsai, B. Potsaid, H. Mashimo, and J. G. Fujimoto, "Correction of rotational distortion for catheter-based en face OCT and OCT angiography," *Opt. Lett.* **39**(20), 5973 (2014).
23. N. Uribe-Patarroyo and B. E. Bouma, "Rotational distortion correction in endoscopic optical coherence tomography based on speckle decorrelation," *Opt. Lett.* **40**(23), 5518 (2015).
24. E. Abouei, A. M. D. Lee, H. Pahlevaninezhad, G. Hohert, M. Cua, P. Lane, S. Lam, and C. MacAulay, "Correction of motion artifacts in endoscopic optical coherence tomography and autofluorescence images based on azimuthal en face image registration," *J. Biomed. Opt.* **23**(01), 1 (2018).

25. G. van Soest, J. G. Bosch, and A. F. W. van der Steen, "Azimuthal registration of image sequences affected by nonuniform rotation distortion," *IEEE Trans. Inform. Technol. Biomed.* **12**(3), 348–355 (2008).
26. T. Wang, C. Lancée, R. Beurskens, J. Meijer, B. Knapen, A. F. W. Van Der Steen, and G. Van Soest, "Development of a high-speed synchronous micro motor and its application in intravascular imaging," *Sens. Actuators, A* **218**, 60–68 (2014).
27. T. Wang, G. van Soest, and A. F. W. van der Steen, "A Micromotor Catheter for Intravascular Optical Coherence Tomography," *eng.* **1**(1), 015–017 (2015).
28. V. Rengarajan, Y. Balaji, and A. N. Rajagopalan, "Unrolling the shutter: CNN to correct motion distortions," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua*, 2345–2353 (2017).
29. A. M. D. Lee, G. Hohert, P. T. Angkiriwang, C. MacAulay, and P. Lane, "Dual-beam manually-actuated distortion-corrected imaging (DMDI) with micromotor catheters," *Opt. Express* **25**(18), 22164 (2017).
30. T. Wang, T. Pfeiffer, E. Regar, W. Wieser, H. van Beusekom, C. T. Lancee, G. Springeling, I. Krabbendam, A. F. W. van der Steen, R. Huber, and G. van Soest, "Heartbeat OCT: in vivo intravascular megahertz-optical coherence tomography," *Biomed. Opt. Express* **6**(12), 5021 (2015).
31. T. S. Kim, J. Y. Joo, I. Shin, P. Shin, W. J. Kang, B. J. Vakoc, and W. Y. Oh, "9.4 MHz A-line rate optical coherence tomography at 1300 nm using a wavelength-swept laser based on stretched-pulse active mode-locking," *Sci. Rep.* **10**(1), 1–9 (2020).
32. S. Tozburun, C. Blatter, M. Siddiqui, E. F. J. Meijer, and B. J. Vakoc, "Phase-stable Doppler OCT at 19 MHz using a stretched-pulse mode-locked laser," *Biomed. Opt. Express* **9**(3), 952 (2018).
33. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. 1981 DARPA Image Underst. Work.* 121–130 (1981).
34. O. O. Ahsen, K. Liang, H.-C. Lee, M. G. Giacomelli, Z. Wang, B. Potsaid, M. Figueiredo, Q. Huang, V. Jayaraman, and J. G. Fujimoto, "Assessment of Barrett's esophagus and dysplasia with ultrahigh-speed volumetric en face and cross-sectional optical coherence tomography," *Endoscopy* **51**(04), 355–359 (2019).
35. I. Sobel, "History and definition of the sobel operator," Retrieved from World Wide Web 1505 (2014).
36. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
37. K. Liang, O. O. Ahsen, A. Murphy, J. Zhang, T. H. Nguyen, B. Potsaid, M. Figueiredo, Q. Huang, H. Mashimo, and J. G. Fujimoto, "Tethered capsule en face optical coherence tomography for imaging Barrett's oesophagus in unsedated patients," *BMJ Open Gastroenterol.* **7**(1), e000444 (2020).
38. M. J. Gora, J. S. Sauk, R. W. Carruth, K. A. Gallagher, M. J. Suter, N. S. Nishioka, L. E. Kava, M. Rosenberg, B. E. Bouma, and G. J. Tearney, "Tethered capsule endomicroscopy enables less-invasive imaging of gastrointestinal tract microstructure," *Nat. Med.* **19**(2), 238–240 (2013).
39. M. J. Gora, L. Quéñéhervé, R. W. Carruth, W. Lu, M. Rosenberg, J. S. Sauk, A. Fasano, G. Y. Lauwers, N. S. Nishioka, and G. J. Tearney, "Tethered capsule endomicroscopy for microscopic imaging of the esophagus, stomach, and duodenum without sedation in humans (with video)," *Gastrointest. Endosc.* **88**(5), 830–840.e3 (2018).
40. T. Houston and P. Sharma, "Volumetric laser endomicroscopy in Barrett's esophagus: Ready for primetime," *Transl. Gastroenterol. Hepatol.* **5**, (2020).
41. H.-C. Lee, O. O. Ahsen, K. Liang, Z. Wang, C. Cleveland, L. Booth, B. Potsaid, V. Jayaraman, A. E. Cable, H. Mashimo, R. Langer, G. Traverso, and J. G. Fujimoto, "Circumferential optical coherence tomography angiography imaging of the swine esophagus using a micromotor balloon catheter," *Biomed. Opt. Express* **7**(8), 2927 (2016).
42. S. Baker, E. Bennett, S. B. Kang, and R. Szeliski, "Removing rolling shutter wobble," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2392–2399 (2010).
43. E. Ringaby and P. E. Forssén, "Efficient video rectification and stabilisation for cell-phones," *Int. J. Comput. Vis.* **96**(3), 335–352 (2012).
44. Y. Lao and O. Ait-Aider, "A Robust Method for Strong Rolling Shutter Effects Correction Using Lines with Automatic Feature Selection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 4795–4803 (2018).
45. K. Liang, G. Traverso, H.-C. Lee, O. O. Ahsen, Z. Wang, B. Potsaid, M. Giacomelli, V. Jayaraman, R. Barman, A. Cable, H. Mashimo, R. Langer, and J. G. Fujimoto, "Ultrahigh speed en face OCT capsule for endoscopic imaging," *Biomed. Opt. Express* **6**(4), 1146 (2015).
46. M. J. Gora, M. J. Suter, G. J. Tearney, and X. Li, "Endoscopic optical coherence tomography: technologies and clinical applications [Invited]," *Biomed. Opt. Express* **8**(5), 2405 (2017).
47. A. López-Marín, G. Springeling, R. Beurskens, H. van Beusekom, A. F. W. van der Steen, A. D. Koch, B. E. Bouma, R. Huber, G. van Soest, and T. Wang, "Motorized capsule for shadow-free OCT imaging and synchronous beam control," *Opt. Lett.* **44**(15), 3641 (2019).
48. I. J. M. Levink, H. C. Wolfsen, P. D. Siersema, M. B. Wallace, and G. J. Tearney, "Measuring Barrett's Epithelial Thickness with Volumetric Laser Endomicroscopy as a Biomarker to Guide Treatment," *Dig. Dis. Sci.* **64**(6), 1579–1587 (2019).
49. O. O. Ahsen, K. Liang, H.-C. Lee, Z. Wang, J. G. Fujimoto, and H. Mashimo, "Assessment of chronic radiation proctopathy and radiofrequency ablation treatment follow-up with optical coherence tomography angiography: A pilot study," *World J. Gastroenterol.* **25**(16), 1997–2009 (2019).