Fast Intelligent Autofocusing Retinal Optical Coherence Tomography for Real-Time Defocus Compensation

Pablo Ortiz
Duke University
Durham, NC, 27708
pablo.ortiz@duke.edu

November 24, 2020

Abstract

Retinal Optical Coherence Tomography [1] is an imaging modality that provides non-invasive *in vivo* volumetric imaging of retinal tissue with micron-scale resolution. OCT combines confocal and coherence gating to provide sectioning of tissue with high axial resolution. Because of this, a defocus caused by refractive error in the subject's eye leads to lower sensitivity due to confocal and coherence gate mismatch. Lateral resolution is determined by the spot size of the beam at the focal plane. This means defocus also leads to lower lateral resolution due to larger spot size. In this paper, we introduce a novel algorithm to based on convolutional neural networks that determines, from a single b-scan, the level of defocus on a system to allow it to correct for it in real time. Our results indicate that the system can correct for any level of defocus after less than 0.3 seconds and without any operator intervention.

1 Introduction

Optical Coherence Tomography (OCT) [2] is a 3D retinal imaging modality that is widespread in the ophthalmic diagnostics field. The most common OCT systems collect data by applying a fourier analysis on interferometric signal captured through a coherent interferometer. This analysis decodes the depth of scatterers along the path of the light source. Retinal OCT systems have a combination of confocal and coherent gating [3]. Confocal gating is present because the collected signal gets collected by an optical fiber tip with very small aperture size that is optically conjugate to the retina. This means that any returning signal not in focus at the fiber tip plane will be coupled less efficiently, as it misses the aperture of the fiber tip. In addition to the confocal gate, OCT has a coherence gate, where a path length difference between the two arms of

the interferometer decreases the coherence between the two interfering beams, leading to a loss of signal of scatterers with a large path length difference. The system is optimized when both gates are centered at the same plane. These systems are typically designed to image subjects whose corneal refractive power exactly focuses a planar wavefront profile at the retinal plane. That way the system can be optimized to match coherence and confocal gates at the retinal plane to maximize the signal collected.

However, the majority of subjects have either some degree of focal refractive error in the cornea, making it such that a planar wavefront does not optically focus exactly at the retinal plane, or an eye length that's different from the average population's eye length, which also results in the optical beam not properly focusing at the retinal plane. Imaging subjects with refractive error leads to two main consequences in the quality of the obtained OCT volumes: 1. a mismatch between confocal and coherent gates that manifests as a loss in sensitivity, and therefore signal-to-noise ratio (SNR) of the system, and 2. an out of focus image that degrades the sharpness of the volume in the lateral direction. The most common way to solve this problem is by designing the OCT scanner such that either the collimating lens or the objective lens are translatable [4] or deformable [5]. That way, the wavefront can be defocused in the opposite way of the eye optics, canceling each other out and refocusing the beam at the retinal plane to recover both SNR and lateral resolution. Usually, an operator is in charge of adjusting this component to maximize image quality. It is of interest to have an automatic and reliable system that can do this quickly and accurately.

2 Related Work

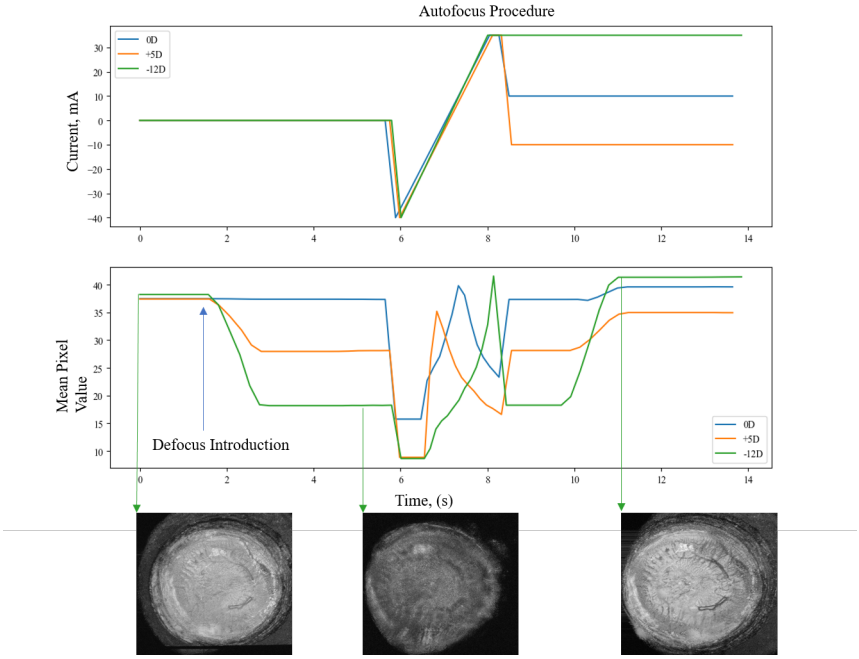

Figure 1: Traditional metric-based autofocusing pipeline. The top plot shows the sweep procedure of the physical property that can correct for focus. In this case it's the current provided to the electrically tunable lens for eyes with different levels of defocus compensation. The central plot shows the average intensity image metric across the sweep and how at the end of it, the system optimized current to match the current that maximized intensity during the sweep. The bottom plot are *en face* projections of the volume acquired at different stages of the procedure.

There is significant work in the field of adaptive optics OCT (AO-OCT) to automate this procedure. In AO-OCT, typically, a wavefront sensor is added to the system to measure the shape of the wavefront returning from the sample and quantitatively measure its deformation. With this information, a deformable mirror is then tuned to apply the opposite deformation and optimize image quality. These systems are very accurate and can correct for optical deformations of higher order than defocus, but they are not currently available widespread due to the increased cost in hardware required to implement it. Without a wavefront sensor automatic focus systems rely on image quality to determine the level of defocus needed to optimize the images.

Wavefront-sensorless autofocusing OCT systems typically aim to maximize

an image metric such as image intensity [4]. To do this, the system sweeps the adjustable optical element and takes an image at each step of the sweep to map the image metric to defocus position (See Figure 1). After that, it picks the level of defocus that maximized the image metric and proceeds to image with that level of defocus. There has been more work to optimize the search strategy and avoid sweeping the entire range and finding the maximum more quickly [6] (2s).

Although these methods of optimizing the image quality have proven to be successful, we aim to improve upon them by providing a machine-learning based algorithm that can, from a single image, determine the level of defocus in the system without having to search the space for the maximization of an image metric. This approach has two main advantages over the ones with a search strategy: 1. they can find the defocus level in a single step, and 2. they don't rely on an image metric, which doesn't have a direct correlation with image quality.

3 Methods

All the imaging was performed with a swept source OCT system shown in Figure 2. For the most part, this is a traditional 4F retinal scanner with the only notable exception being the tunable lens (Optotune A.G., Switzerland) adjacent to the collimator that provides the focus adjustment capability. The tunable lens has a diopter correction range of -13 to 13 diopter, which, in our system, is able to correct for defocus from the corneal surface on the range of -10 to 10 diopters without losing more than 40% sensitivity at the extrema.

Even though OCT is a volumetric imaging technique, we trained our model on 1376x500 cross-sectional b-scan images because they are acquired at a much faster rate (8ms) than that of the volumes (1s), but still provide enough image quality information for autofocusing.

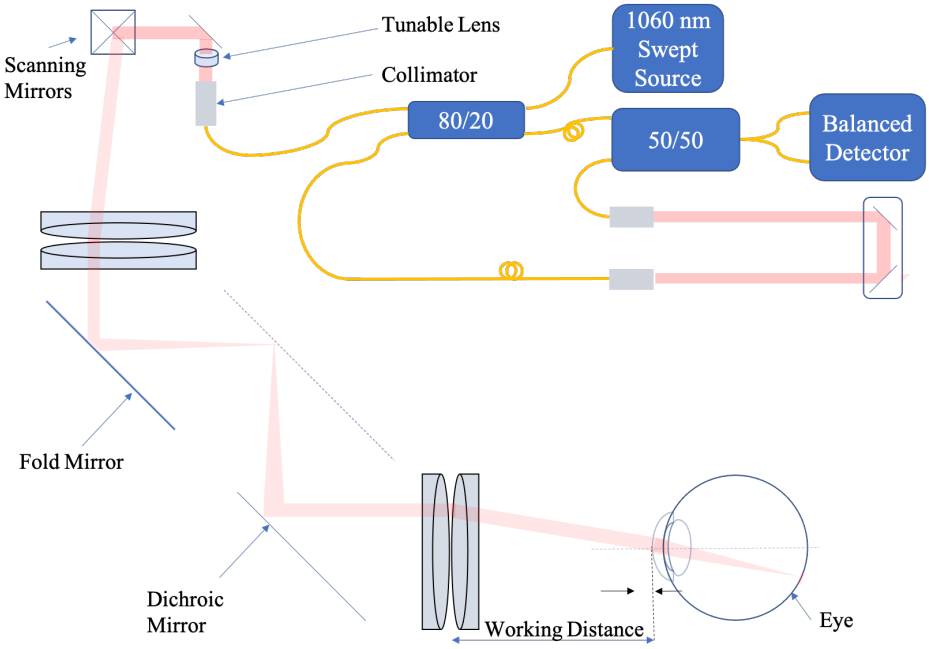

Figure 2: Imaging System with a 4F telescope relaying scanning mirrors to the pupil plane and a tunable lens for defocus compensation next to the optical collimator used to couple light into and out of the optical fibers. The shape of the tunable lens introduces a wavefront aberration at a pupil-conjugate plane that compensates for that caused by the cornea.

In order to obtain a dataset with images required for training a network that can determine the level of defocus from a single b-scan, we imaged a retinal eye model (Rowe Technical Design, CA). We decided to capture images instead of modelling them because of the complex nature of OCT imaging, among the main reasons: axial resolution and lateral resolution are decoupled; online datasets captured in focus are still not in focus across the entire volume as the depth of focus is narrow in OCT, so they are not a valid ground truth model; most datasets available online do not account for index of refraction deformation caused by the curvature of the cornea, so images are distorted compared to the real retina; the combination of confocal and coherence gating is non-trivial to simulate and varies across depth planes through the volume; and finally, the 3D nature introduces artifacts that are difficult to model such as multiple-scattering and speckle. All of these factors make it so that it's easier and more accurate to capture real data than it is to attempt to simulate it.

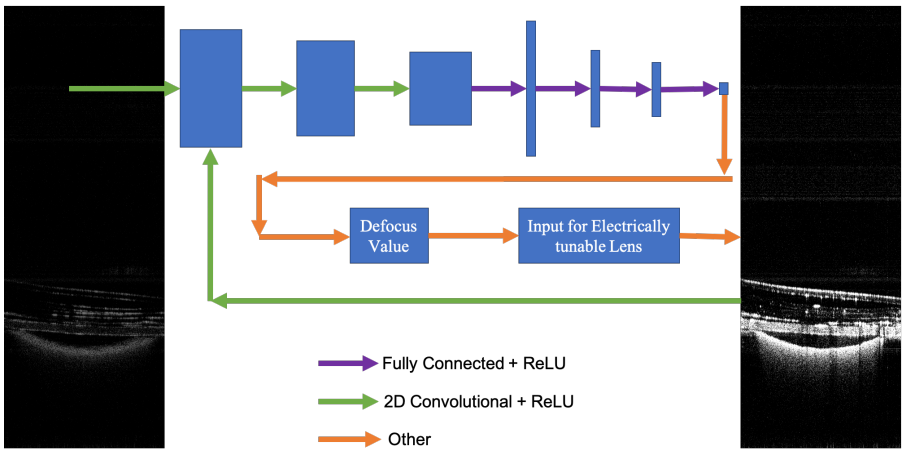

Figure 3: Pipeline for real-time autofocusing. The first row of the pipeline represents the neural network. The filter size/number of filters/stride for the three convolutional layers are 8x8/4/8, 4x4/8/4, and 4x4/8/4. The dimensions of the fully-connected layers are 1024, 512, 10, and 1. The input size of the bscan is 1376 x 500.

3.1 Data Acquisition

To create our dataset, we manually optimized the image quality by adjusting the focal power of the tunable lens until the image was as crisp and bright as possible. Once the image was optimized, we recorded the current level provided to the electrically tunable lens (6mA, in our case) and stored that number as the optimal current value. We then introduced defocus by applying different currents to the tunable lens, from -30mA to 30mA at 1mA increments and imaged the model eye, storing 40 b-scans evenly distributed across the volume at each defocus current. The label for each of those b-scans was the level of defocus, evaluated as the absolute value of the difference between the applied current and the optimal current value. This amounted for a total of 2400 b-scans: 60 defocus levels and 40-bscans per defocus level. This data was later randomly split 80:20 into training:validation datasets, respectively.

3.2 Training Details

The network, which was adapted from a digital camera autofocusing neural network [7], is shown in the top row of Figure 3. The network was trained using TensorFlow on a Google Colab GPU runtime. The loss function was the mean squared error and the model was trained with an Adam optimizer. We used a batch size of 32 and a learning rate of 0.001 over 300 epochs.

3.3 Model integration

Once the model was trained, the weights were stored and downloaded to the OCT computer. The model was recreated in the OCT acquisition pipeline and the weights were loaded. Every 100ms, the OCT acquisition provided the last acquired b-scan into the neural network to evaluate the level of defocus. It then automatically adjusted the focus level of the electrically tunable lens. The system randomly picked on direction to move in the first step. For subsequent steps, the choice between forward and backward was made by choosing the step that was closest to previous evaluated candidates.

3.4 Performance Assessment

To evaluate the quality of our system and its effectiveness in quickly autofocusing images in real-time, we imaged the model eye and introduced a large defocus, unknown to the autofocus algorithm, and waited for the system to recover focus. We estimated the amount of time it took to recover focus by calculating the percentage of the volume the image remained out of focus given that the volume acquisition rate is 1 volume every 1.45 seconds 4. Because our autofocus implementation is written in Python, but the acquisition software is written in C++, there is some overhead to provide the b-scan data to the neural network. This means that, even though we acquire b-scans at a rate of 8ms per b-scan, the control cycle of our autofocus algorithm happens once every 100ms. We then measured the number of control cycles it takes our system to find the correct level of focus.

4 Results

The training loss of the model after training turned out 0.0456 with a validation loss of 0.0192. For context, the units represent current, in mA, of defocus measure. The tunable lens has a digital control of 0.7mA steps. This means that the accuracy of the training model is an order of magnitude more precise than the hardware limit we have to control for focus and therefore more than sufficient for our application.

Qualitatively, images were observed to recover focus as observed by both the intensity and sharpness of the *en face* projections shown in Figure 4. The total time needed for autofocusing was measured to be in the order of 200m, even for defocus introductions that are much larger than those be expected from human subject's refractive error. Figure 4, for example, has a defocus level that completely removes the image acquired, as seen in the dark band, which wouldn't happen even in heavily miopic subjects.

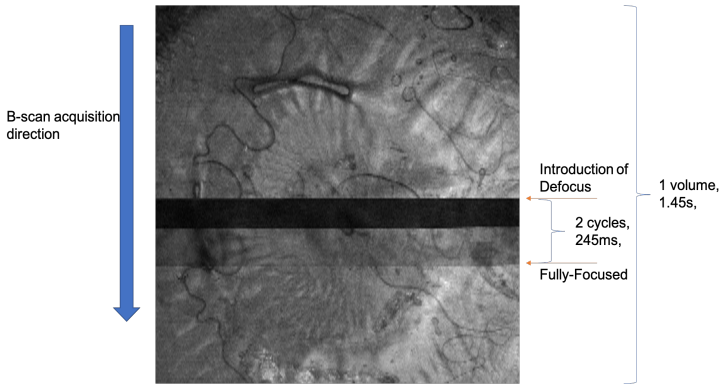

Figure 4: *En face* summed volume projection of a volume acquired with a sudden defocus introduced halfway through acquisition. The dark band shown was the time period over which the system was fully out of focus. After 1 control cycle of 120ms, the image recovers most of the focus, and after another control cycle of 125ms, it recovers focus entirely.

5 Discussion

This methodology for automatic focusing in OCT demonstrates to be not only faster than the state-of-the-arts focus speeds (roughly 3s) [8], but has the additional advantage that it can be constantly being performed during an imaging session as opposed to triggered by an operator only once and assuming the focus level does not change during the acquisition. It also has the advantage that it does not require expensive hardware to be implemented (the tunable lens is only $125).

For this methodology to be adopted and useful, however, it is of crucial importance to expand the training dataset past a retinal model and acquire b-scans from human subjects, as well as demonstrate the performance is just as good on human subjects as it is with the model. However, it is worth mentioning that the results of this paper are still of importance even with a single imaging model for data. That is because b-scans of human retina do not differ significantly across subjects. In particular, the reflectivity of the retinal layers and the thickness of it, which are the important features that defocus affects, are relatively stable across subjects, so we can expect a machine learning based automatic focusing to work robustly across the population with a reasonably small training set.

Overall, the results in this paper demonstrate big potential in machine learning based fast autofocusing, but require demonstration on human *in vivo* tissue before it can be widely adopted.

References

[1] Eric A Swanson, Joseph A Izatt, Michael R Hee, David Huang, CP Lin, JS Schuman, CA Puliafito, and James G Fujimoto. In vivo retinal imaging by optical coherence tomography. *Optics letters*, 18(21):1864–1866, 1993.

[2] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and et al. Optical coherence tomography. *Science (New York, N.Y.)*, 254(5035):1178–1181, 1991.

[3] Aaron D. Aguirre, Juergen Sawinski, Shu-Wei Huang, Chao Zhou, Winfried Denk, and James G. Fujimoto. High speed optical coherence microscopy with autofocus adjustment and a miniaturized endoscopic imaging probe. *Opt. Express*, 18(5):4222–4239, Mar 2010.

[4] Maddipatla Reddikumar, Ayano Tanabe, Nobuyuki Hashimoto, and Barry Cense. Optical coherence tomography with a 2.8-mm beam diameter and sensorless defocus and astigmatism correction. *Journal of biomedical optics*, 22(2):026005, 2017.

[5] Yuankai K. Tao, Sunil K. Srivastava, and Justis P. Ehlers. Microscope-integrated intraoperative oct with electrically tunable focus and heads-up display for imaging of ophthalmic surgical maneuvers. *Biomed. Opt. Express*, 5(6):1877–1885, Jun 2014.

[6] Yifan Jian, Jing Xu, Martin A Gradowski, Stefano Bonora, Robert J Zawadzki, and Marinko V Sarunic. Wavefront sensorless adaptive optics optical coherence tomography for in vivo retinal imaging in mice. *Biomedical optics express*, 5(2):547–559, 2014.

[7] Chengyu Wang, Qian Huang, Ming Cheng, Zhan Ma, and David J. Brady. Intelligent autofocus, 2020.

[8] Hans RGW Verstraete, Morgan Heisler, Myeong Jin Ju, Daniel Wahl, Laurens Bliek, Jeroen Kalkman, Stefano Bonora, Yifan Jian, Michel Verhaegen, and Marinko V Sarunic. Wavefront sensorless adaptive optics oct with the done algorithm for in vivo human retinal imaging. *Biomedical Optics Express*, 8(4):2261–2275, 2017.