tvst

DOI: 10.1167/tvst.5.1.1

Article

Virtual Averaging Making Nonframe-Averaged Optical Coherence Tomography Images Comparable to Frame-Averaged Images

Chieh-Li Chen[1,2], Hiroshi Ishikawa[1,2], Gadi Wollstein[1,2], Richard A. Bilonick[1,3], Larry Kagemann[1,2], and Joel S. Schuman[1,2,4]

[1] UPMC Eye Center, Eye and Ear Institute, Ophthalmology and Visual Science Research Center, Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[2] Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, USA
[3] Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
[4] McGowan Institute for Regenerative Science, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Corresponding Author: Gadi Wollstein, MD, UPMC Eye Center, Eye and Ear Institute, 203 Lothrop Street, Pittsburgh, PA 15213; e-mail: wollsteing@upmc.edu

Received: 29 June 2015
Accepted: 29 November 2015
Published: 12 January 2016

Keywords: image processing; optical coherence tomography; retina

Citation: Chen C-L, Ishikawa H, Wollstein G, Bilonick RA, Kagemann L, Schuman JS. Virtual averaging making nonframe-averaged optical coherence tomography images comparable to frame-averaged images. 2016;5(1):1, doi:10.1167/tvst.5.1.1

Purpose: Developing a novel image enhancement method so that nonframe-averaged optical coherence tomography (OCT) images become comparable to active eye-tracking frame-averaged OCT images.

Methods: Twenty-one eyes of 21 healthy volunteers were scanned with noneye-tracking nonframe-averaged OCT device and active eye-tracking frame-averaged OCT device. Virtual averaging was applied to nonframe-averaged images with voxel resampling and adding amplitude deviation with 15-time repetitions. Signal-to-noise (SNR), contrast-to-noise ratios (CNR), and the distance between the end of visible nasal retinal nerve fiber layer (RNFL) and the foveola were assessed to evaluate the image enhancement effect and retinal layer visibility. Retinal thicknesses before and after processing were also measured.

Results: All virtual-averaged nonframe-averaged images showed notable improvement and clear resemblance to active eye-tracking frame-averaged images. Signal-to-noise and CNR were significantly improved (SNR: 30.5 vs. 47.6 dB, CNR: 4.4 vs. 6.4 dB, original versus processed, $P < 0.0001$, paired *t*-test). The distance between the end of visible nasal RNFL and the foveola was significantly different before (681.4 vs. 446.5 μm, Cirrus versus Spectralis, $P < 0.0001$) but not after processing (442.9 vs. 446.5 μm, $P = 0.76$). Sectoral macular total retinal and circumpapillary RNFL thicknesses showed systematic differences between Cirrus and Spectralis that became not significant after processing.

Conclusion: The virtual averaging method successfully improved nontracking nonframe-averaged OCT image quality and made the images comparable to active eye-tracking frame-averaged OCT images.

Translational Relevance: Virtual averaging may enable detailed retinal structure studies on images acquired using a mixture of nonframe-averaged and frame-averaged OCT devices without concerning about systematic differences in both qualitative and quantitative aspects.

Introduction

Since its first appearance in the clinical application, optical coherence tomography (OCT) has become a standard in daily clinical care in ophthalmology. The technology provides high-resolution cross-sectional images in a noncontact and noninvasive fashion as well as objective assessment of the tissue structures.[1–8] Currently, multiple spectral-domain (SD-) OCT devices from several manufacturers are commercially available. While all manufacturers aim to improve OCT technology in order to improve their clinical diagnoses, disease detection ability, measurement

reproducibility, signal quality, and image appearance, each manufacturer also dedicates their efforts to advancing their products without paying much attention to data/measurement compatibility and comparability. This diversity causes OCT data inconsistency and OCT measurement incompatibility,[9–12] and further limits the application of OCT in the clinical practice for clinicians need to stick to one OCT device, to assure the measured thickness data are consistent and comparable for their follow-up patients.[10–17]

Among all the advances in both hardware and software, eye-tracking system has been recognized as one of the most valuable developments in SD-OCT device lately. Active eye-tracking systems in OCT devices are used for two purposes: (1) frame averaging to improve signal-to-noise ratio (SNR), and (2) ensuring consistent image registration from visit to visit. Although this is a great technological advancement, it creates even wider diversity and inconsistency among OCT devices in terms of signal quality, image appearance, and quantitative measurements, especially in multicenter studies, where each center collects OCT data with different devices, or in longitudinal studies, where the devices used to acquire image data changed with time.

Previously, we developed a signal normalization method, which successfully normalized OCT signal characteristics and reduced the measurement differences between two nonframe-averaged SD-OCT devices.[18,19] Yet, this method did not work well in reducing the retinal thickness measurement differences between nonframe-averaged and frame-averaged images (unpublished data). To solve this problem, we developed a method called virtual averaging. We hypothesize that both signal characteristics variation and the systematic differences in OCT measurements between nonframe-averaged and frame-averaged OCT images get reduced by virtually averaging nonframe-averaged OCT signal and integrating it into our signal normalization method.

Methods

Subjects and Image Acquisitions

Twenty-one healthy subjects volunteered to participate in this prospective cross-sectional study. The right eye from each subject was used in the study. All subjects were recruited at the University of Pittsburgh Medical Center Eye Center. The University of Pittsburgh institutional review board and ethics committee approval were obtained for the study, and informed consent was obtained from all subjects. This study followed the tenets of the Declaration of Helsinki and was conducted in compliance with the Health Insurance Portability and Accountability Act.

For each participant, both macular and optic nerve head (ONH) regions were scanned at the same visit using two commercially available SD-OCT devices, where one generated nonframe-averaged images while the other had eye-tracking system and provided frame-averaged images.

Nonframe-Averaged OCT Image Acquisition

Cirrus HD-OCT (software version 6.1; Zeiss, Dublin, CA) was used to collect nonframe-averaged OCT image data, where Macula Cube 200×200 and Optic Disc 200×200 scan protocols were used. These scans consisted of 200×200 sampling points on a 6×6 mm^2 area centered on the foveola and the center of ONH, respectively, and 1024 samplings within a 2.0-mm axial scan depth.

Frame-Averaged OCT Image Acquisition

Spectralis (software version 1.5; Heidelberg Engineering, Heidelberg, Germany) was used to collect corresponding active eye-tracking frame-averaged OCT image data. Macular raster volume scan centered at the foveola covering $20° \times 20°$ region (193 sections, 1024 sampling points in one section, 9-frame averaged), and Circle retinal nerve fiber layer (RNFL) scan (circumpapillary RNFL or cpRNFL; 100-frame averaged) were used to acquire the image data for macular and ONH regions, where 496 sampling points were collected within 1.9-mm axial scan depth.

For both Cirrus and Spectralis data, multiple scans were acquired at the same visit to ensure that the collected OCT images have good signal quality and spatial integrity for further processing. Images with image quality below the manufacturer recommended cutoff, signal strength (SS) below 6 for Cirrus images and image quality score below 15 for Spectralis images, retina structures outside the scanning window, or images with apparent eye movement during scanning were considered poor quality images and discarded. Eye movement was subjectively defined as image artifacts on OCT *en face* (or OCT fundus) images showing a horizontal frame shift larger than one average size retinal blood vessel diameter and a major distortion of the fovea or ONH region. All the image raw data files (for Cirrus data, we used the z-

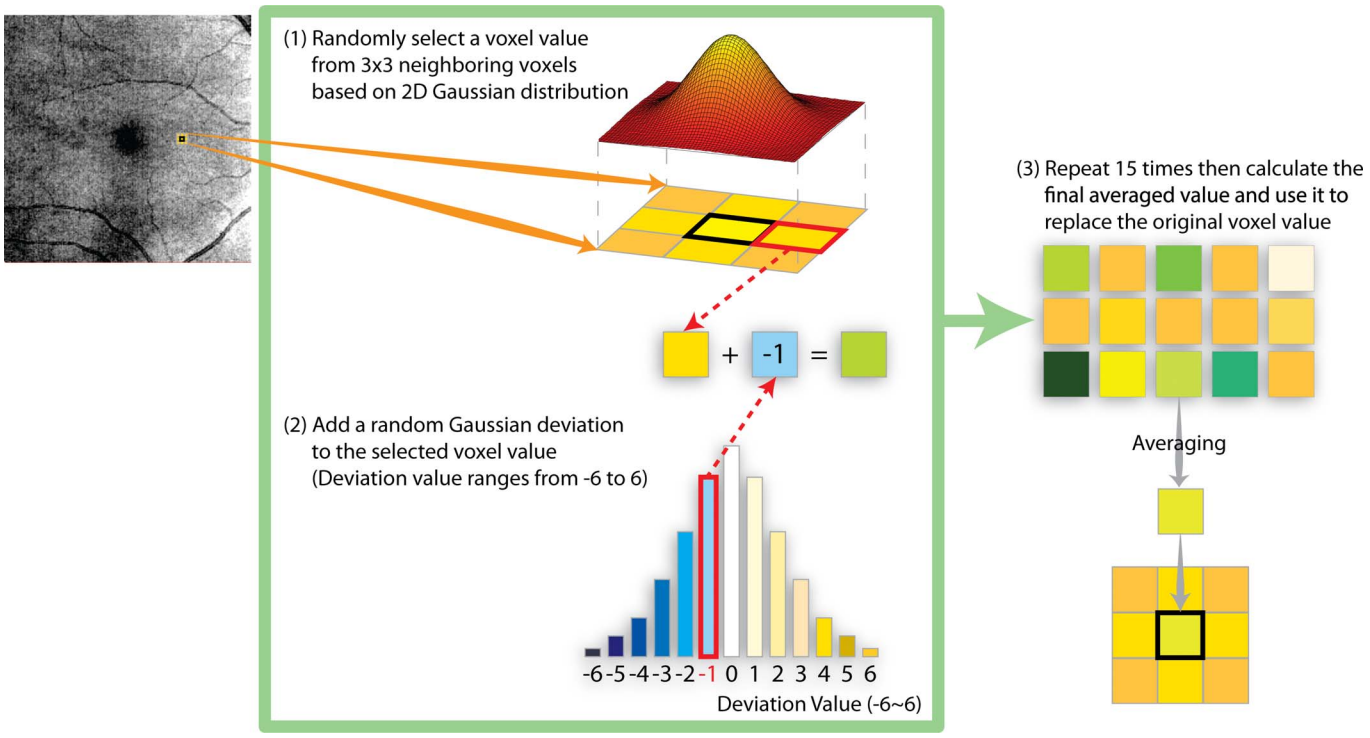

Figure 1. Processing flow of virtual averaging. (1) For each sampling voxel (the *center square* in Step 1, with *thick black border*), one neighboring voxel out of nine (including the center) on the same *z*-position was randomly selected according to a 2D Gaussian distribution. (2) A random Gaussian deviation was added to the selected voxel value. (3) Repeating previous two steps 15 times, and the new voxel value was calculated by averaging all 15 values replacing the original voxel value.

motion corrected raw data) were exported to a standalone computer for signal normalization and further processing and analysis.

Virtual Averaging

To bridge the gap between nonframe-averaged and frame-averaged OCT signal characteristics, we developed a method called virtual averaging, and applied this method to nonframe-averaged OCT images.

The eye-tracking Spectralis scanner scans at the same location multiple times to generate frame-averaged OCT data. Even with the active eye-tracking system, scanning location deviates every time when the scanner tries to scan the same location as fixational eye movement is faster than tracking rate.[20–22] We observed that every time the scanner locates the camera back to the same location, it results in the acquisition deviation in both scanning location and signal strength (mostly due to fixational eye movement). Therefore, in order to mimic both the acquisition deviation, we first simulated the deviations in the x (transverse) and y (vertical) directions. For each sampling voxel, one neighboring voxel was randomly selected from the 3×3 neighboring voxels located on the same z-position (including the center voxel) following a two-dimensional (2D) Gaussian random distribution, where the closer the voxel was to the center, the higher its likelihood of being selected as a candidate (as Fig. 1, Step 1 shows). Then we simulated the variation of the voxel value by adding a random Gaussian deviation to the selected voxel value, where the range of the voxel value deviation was from -6 to 6 (Fig. 1, Step 2). This process (deviations in x and y directions and in voxel value, namely Steps 1 and 2 in Fig. 1) was repeated 15 times for each voxel, and the average of the outcomes were used to replace the original value (Fig. 1, Step 3).

All the parameters used in virtual averaging processing were determined based on our internal testing (unpublished data), observations, and the goal of not to generate much blurring artifacts in the retinal tissues. Considering the coarser sampling density in both x and y directions of Cirrus device, we set the kernel size for sampling deviation to $3 \times 3 \times 1$ ($x \times y \times z$) so that the resampled A-scans would not have too much variation, especially near the foveola. The range of the amplitude deviation (ranged from -6 to 6) was decided based on the observation of pixel

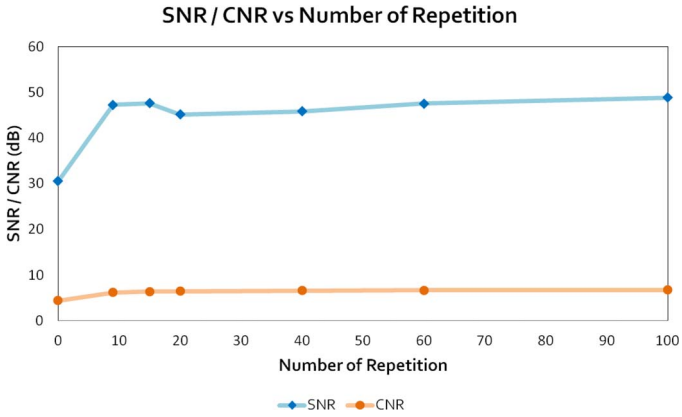

Figure 2. Plot of SNR and CNR change as the number of repetition in virtual averaging processing increases. Signal-to-noise ratio reached a local maximum while CNR reach a plateau when the number of repetition was at 15.

intensity variation in both Cirrus and Spectralis data and our preliminary experiments. A wide range of the deviation value was tested. The sigma of the amplitude deviation Gaussian distribution was set to 1, 2, 3, 5, 7, and 10, which covered a wide range of the actual intensity deviation in both Cirrus and Spectralis data. When the sigma is too small, like 1 and 2, the signal in each retinal layer looks granular; on the other hand, when the sigma is too big, for example, 7 and 10, speckle-noise-like artifacts appear. Taking all the concerns into account, the results with a sigma of 3 showed clearer and smoother cross-sectional images compared with other settings. Therefore for amplitude deviation, we chose the sigma to be 3, which led to the amplitude deviation ranged from −6 to 6.

We have further tested the virtual averaging results with different repetition times: 9 (the default number for frame averaging processing in Spectralis), 15, 20, 40, 60, and 100 (the maximum number for frame averaging in Spectralis). Pemp et al.[23] showed that averaging with more frames improves the image quality and visualization.[24] However, in our results, the SNR reached a local maximum while contrast-to-noise ratio (CNR) reached a plateau when the number of averaging was at 15 and did not show significant improvements when the number of averaging was larger than or equal to 20 (Fig. 2). Furthermore, the running time required to finishing the virtual averaging processing increased linearly with respect to the repetition times. It took 60.7 seconds for 15-time averaging, 80.4 seconds for 20-time averaging, 159.6 seconds for 40-time averaging, and 400.9 seconds for 100-time averaging (assessed using a MacBook Pro with 2.6 GHz dual-core Intel Core i5 processor; Cupertino, CA). To strike a balance between the outcome image quality and running time, 15 repetitions of Steps 1 and 2 was used in our virtual averaging processing.

Virtual Averaging Performance Assessment

Experiment I: Effects of Image Quality Enhancement

This experiment was designed to test the effects of the proposed virtual averaging method on OCT image quality enhancement, especially focused on the visibility and clarity of the borders that delineate intraretinal layers. To compare and test the image enhancement ability, nonframe-averaged OCT data were also processed with a standard Gaussian smoothing method, a widely known image-smoothing filter for noise reduction and image content enhancement, however, losing image details, and blurring the boundaries.[25] The standard Gaussian smoothing method used in this comparison has the same kernel size as the virtual averaging (i.e., $3 \times 3 \times 1$ in the x, y, and z directions, respectively). The outcomes were evaluated and compared with frame-averaged OCT data subjectively and objectively.

Subjective and objective assessments. For subjective assessment, subjective image quality evaluation based on the image appearance in terms of tissue contrast, smoothness of the tissues, and the visibility of intraretinal layers were assessed independently by two observers (CLC and HI). For objective assessment, SNR and CNR were calculated to evaluate the image enhancement effect as follows[26–30]:

$$\mathrm{SNR} = 10\log\left(\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\max(I^2)}{\sigma_{\mathrm{n}}^2}\right)_m\right),$$

$$\mathrm{CNR} = 10\log\left(\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\mu_{\mathrm{f}}-\mu_{\mathrm{n}}}{\sqrt{\sigma_{\mathrm{f}}^2+\sigma_{\mathrm{n}}^2}}\right)_m^2}\right) \quad (1)$$

In the equation, intermediate SNR and CNR for each frame were calculated and accumulated, then the final SNR and CNR values were assessed using the arithmetic average of the intermediate parameters, where M stands for the number of frames in one cube data. In the expression for SNR, I represents the logged value from the OCT machine output, and σ_{n}^2 stands for the variance of the background noise region in the logged value. In the CNR formula, μ_{f} and μ_{n} indicate the mean value of the selected regions of interest and of the same background noise region

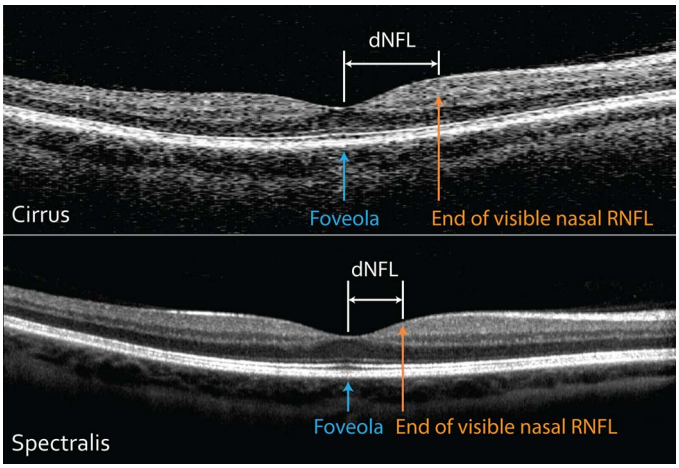

Figure 3. Distance between the end of visible nasal RNFL and the foveola. *Blue arrow* points to the foveola position while *orange arrow* points out the end of visible nasal RNFL. The distance between the end of visible nasal RNFL and the foveola is calculated as the horizontal distance between the *blue* and *orange arrows*.

as in SNR, while σ_f^2 and σ_n^2 stand for the variance of the selected regions of interest and of the same background noise region as in SNR.[26–30]

In addition to the conventional image quality metrics (SNR and CNR), distance between the end of visible nasal RNFL and the foveola (dNFL) was measured to further assess effect of virtual averaging on improving the visibility and clarity of the borders that delineate intraretinal layers. To measure dNFL (done by CLC), the foveola location was manually identified by detecting the largest separation between the junction of the inner and outer segments of the photoreceptors and retinal pigment epithelium (RPE) as appearing on the horizontal and vertical cross-sectional B-scans. On the same frame where the foveola located, the end of the visible nasal RNFL was then manually identified by detecting the largest drop in signal intensity from the nasal side RNFL toward the foveola. The distance between the foveola and the marked A-scan was converted into physical distance in micrometers, and denoted as dNFL. Examples of dNFLs in Cirrus and Spectralis images are presented in Figure 3. Three dNFL measurements were assessed for each image data to evaluate its intraobserver repeatability. In addition, two independent operators conducted the dNFL measurements in a masked manner to test its interobserver reproducibility.

Statistical analysis. Paired *t*-tests were used to analyze the image quality metrics (SNR and CNR) improvement between the original and processed images (both virtual averaged and Gaussian smoothed images) as well as the differences in dNFL between nonframe-averaged and frame-averaged data before and after processing. The intraobserver repeatability and interobserver reproducibility of dNFL measurement were evaluated by calculating the coefficients of variation (CVs) for original image data and frame-averaged data.[31]

Experiment II: Effects of Signal Normalization and Minimizing Tissue Thickness Measurement Differences

Systematic differences exist in the retinal tissue measurements among multiple OCT devices. We hypothesized that the systematic differences are due to the variation in OCT signal characteristics generated from different OCT devices. To solve this problem, in the second experiment, we integrated the virtual averaging technique into our signal normalization and expected that the signal characteristics variation would be minimized after signal normalization, and therefore, could further reduced the measurement differences between nonframe-averaged and frame-averaged images. To test our hypotheses, the ocular tissue thickness measurement differences between Cirrus and Spectralis were measured before and after signal normalization with virtual averaging.

Signal normalization. The signal normalization was modified from the previously reported method.[18,19] In this experiment, Cirrus OCT data format was chosen to be used as the normalization data format reference, and therefore Spectralis OCT data format was converted to the Cirrus equivalent OCT data format. The normalization process had two separate stages for Cirrus and Spectralis data, as presented in Figure 4. Spectralis data were first processed with Z-scaling and sampling density normalization, while Cirrus data were processed with virtual averaging as the first step. Because the averaging of multiple frames during image acquisition for Spectralis data and virtual averaging applied on the Cirrus data have the same effect on reducing the speckle noise, the speckle noise reduction step in the original signal normalization method was omitted. After the first step, both Cirrus and Spectralis data were processed with amplitude normalization to match the intensity dynamic range.

OCT thickness measurement comparison. The thickness measurement comparison methods were the same as the one described in a previous signal normalization study.[19] Briefly, with the assumption that

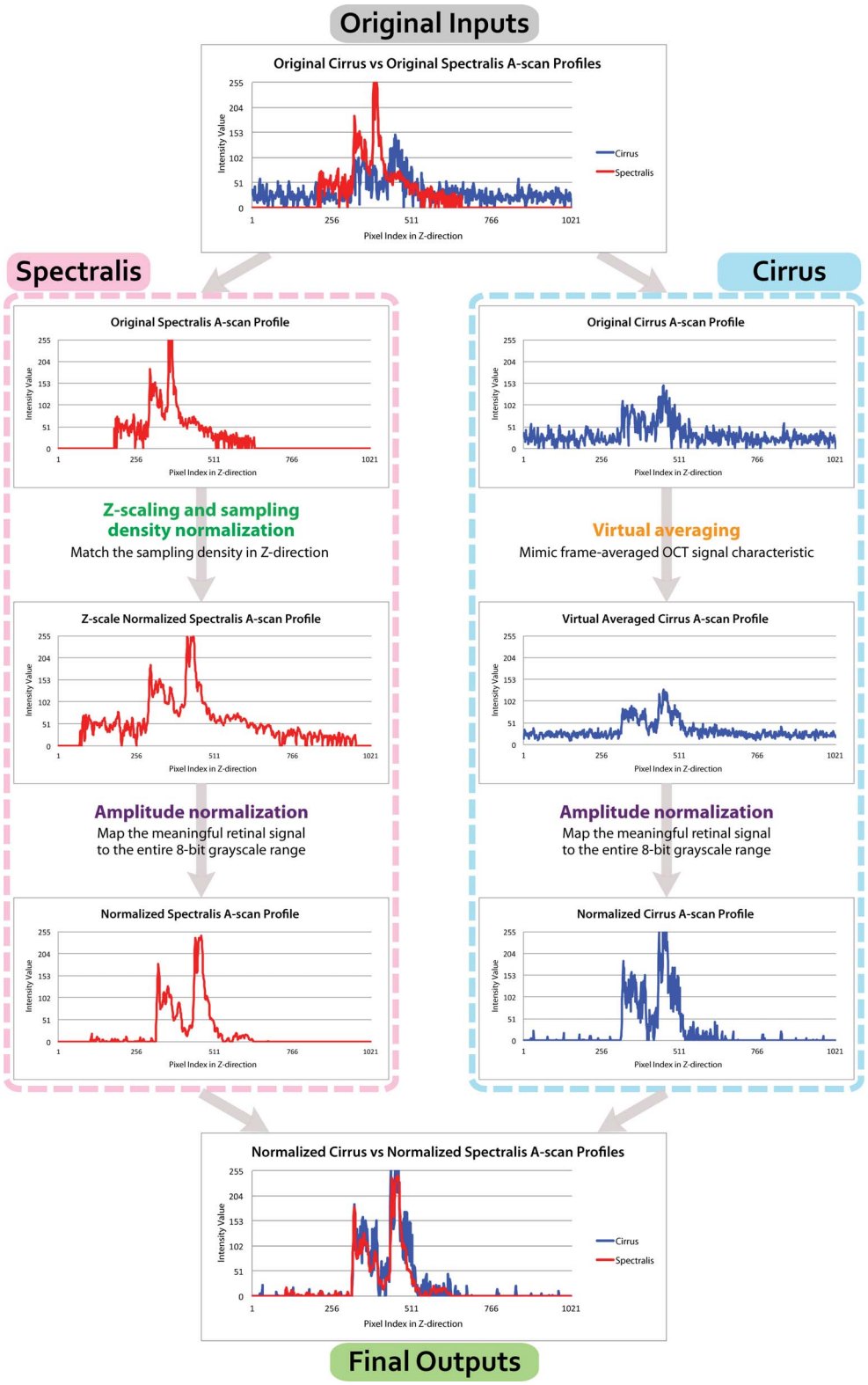

Figure 4. Flow chart of signal normalization with virtual averaging. The A-scan profiles after each step are shown for both Cirrus and Spectralis. Virtual averaging was applied to Cirrus data to mimic frame-averaged signal characteristic, and Z-scaling and sampling density normalization was applied to Spectralis so that the sampling density in Z-direction can be matched to Cirrus specification. Then the amplitude normalization was applied to both processed Cirrus and Spectralis data to maximize the data dynamic range by mapping the meaningful retinal signal to full 8-bit grayscale range. (For display purpose, the original and Z-scale normalized Spectralis A-scan profiles were linearly rescaled from 16-bit to 8-bit grayscale range by mapping the minimal and maximal value to 0 and 255.)

Table 1. Definition of Different Methods for Comparison of RNFL Thickness Measurements

Comparison Methods	OCT Signal	Measurement Description
Comparison I	Original signal	Original device outputs
Comparison II	Original signal	Self-designed segmentation algorithm
Comparison III	Original signal	Self-designed segmentation algorithm with internal parameters optimized for Cirrus and Spectralis individually
Comparison IV	Normalized signal	Self-designed segmentation algorithm

different segmentation algorithms used for ocular tissue thickness measurement in Cirrus and Spectralis devices resulted in the systematic measurement differences, we tested if using the same segmentation algorithm could reduce the systematic measurement differences between OCT devices in addition to the virtual averaging effects on the measurements. Original machine outputs of nine sectoral macular total retinal thicknesses (following an Early Treatment Diabetic Retinopathy Study [ETDRS] pattern[31]) and global mean cpRNFL thickness on the original Cirrus and Spectralis data were collected from the commercial devices and used as the baseline measurements (Comparison I, Table 1). In order to match the measured region for sectoral macular total retinal thickness, the foveola position was manually selected by looking for the largest separation between the junction of inner and outer segment of the photoreceptors and the RPE as appearing on the horizontal and vertical cross-sectional B-scans. Same set of measurements (the nine sectoral macular total retinal thickness and cpRNFL) was automatically measured for both original Cirrus and Spectralis OCT data by applying identical segmentation algorithm of our own design (Comparison II).[32] Then we optimized the internal parameters of our segmentation algorithm for Cirrus and Spectralis data individually (but the core segmentation algorithm was the same) to see if optimizing internal parameters helps minimize the systematic measurement differences (Comparison III). Finally, OCT data were processed with signal normalization including the virtual averaging processing then segmented using our segmentation algorithm without individual internal parameter optimization (Comparison IV).

Image data were excluded if the images demonstrated one or both of the following: (1) apparent inaccurate border detection for more than consecutive 15% or additive 20% of the total image, or (2) borders of the RNFL collapsed, meaning that the RNFL thickness was recorded as a string of zeros for at least 10 consecutive sampling points (the second criteria only applied for cpRNFL measurements).

Statistical analysis. To analyze the absolute difference in nine sectoral total retinal thicknesses in the macular region and cpRNFL thicknesses between Cirrus and Spectralis from four comparison methods, we constructed a comprehensive measurement error model.[19] This measurement error model describes how the true unknown retinal tissue thickness of each eye is linked to the measurements from each device and processing method between Cirrus and Spectralis for different comparison methods.

Structural equation models (SEMs) were used to estimate the parameters in the measurement error model. The R Environment and Language for Statistics (version 2.13.1)[33] with Openx (version 1.1.2-1818),[34] and merror (version 1.0)[35] were used to describe the SEMs. Full information maximum likelihood (FIML) was used to estimate the measurement error model parameters.

Results

Twenty-one healthy subjects, consisted of 9 men and 12 women, were enrolled to this study. The mean age of the cohort was 34.3 ± 11.5-years old. The averaged standard automated perimetery mean deviation (MD) was −0.6 ± 1.1 dB.

Experiment I: Effects of Image Quality Enhancement

Figure 5 presents an example result of the cross-sectional images in the macular region from the original nonframe-averaged Cirrus image (Fig. 5A), Gaussian smoothed Cirrus image (Fig. 5B), virtual averaged Cirrus image (Fig. 5C), and frame-averaged Spectralis image (Fig. 5D). Gaussian smoothed images showed improved image quality compared with the original nonframe-averaged Cirrus images. However, comparing with virtual averaged images, Gaussian smoothed images were blurrier, and showed

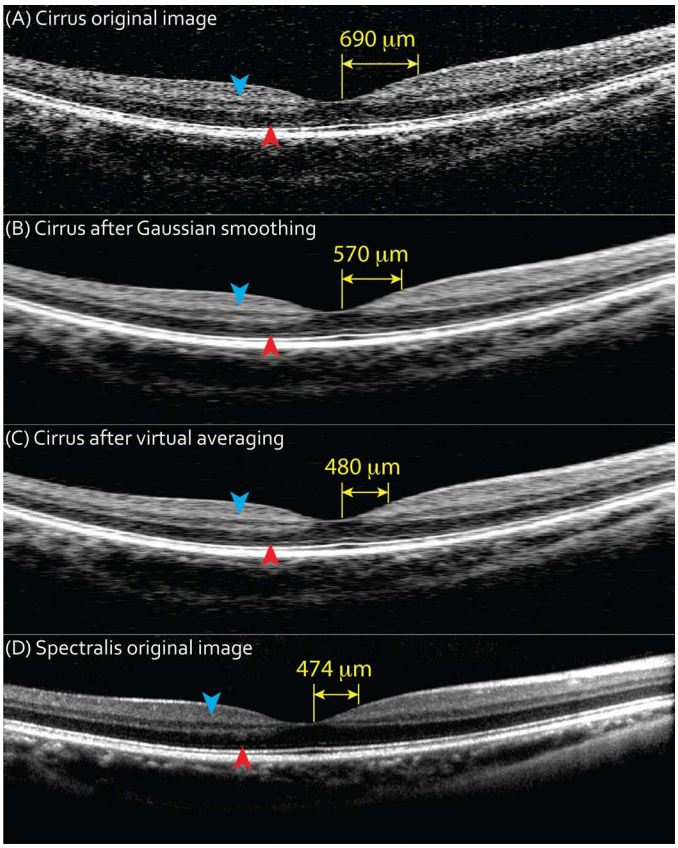

Figure 5. Example results: cross-sectional images of (A) original Cirrus, (B) Cirrus image after Gaussian smoothing, (C) Cirrus image after virtual averaging, and (D) original Spectralis image. Gaussian smoothed image (B) shows improved image quality comparing to original Cirrus image (A), but with blurrier retinal signals and less contrasts between retinal layers than Cirrus image after virtual averaging (C). After virtual averaging, Cirrus image (C) show notable improvement in signal quality and the visibility of retinal tissue than (A). The retina cross-sectional image after virtual averaging looks similar to active tracking frame-averaged Spectralis image (D). In the averaged Cirrus image, external limiting membrane (*red arrowhead*) and the continuous inner border of inner plexiform layer (*blue arrowhead*) become clearly visible.

less signal contrasts between retinal layers. Subjectively, after virtual averaging, all nonframe-averaged images showed notable improvement in image quality and bore clear resemblance to active tracking frame-averaged Spectralis images. As the example shows, the external limiting membrane (ELM; red arrowhead) was hard to recognize in Cirrus original image but became clearly visible and easy to trace after processing. Moreover, the contrast between retinal layers became more apparent and the continuous inner border of inner plexiform layer (IPL; blue arrowhead) became easily distinguishable after virtual averaging.

Table 2. Summarization of the Objective Assessment Results

	SNR (dB)	CNR (dB)
Original	30.5 (28.6, 32.4)	4.4 (4.2, 4.6)
Virtual averaging	47.6 (45.6, 49.5)	6.4 (6.2, 6.7)
Gaussian smoothing	33.0 (31.1, 34.9)	6.7 (6.5, 7.0)

Signal-to-noise ratio (SNR) and CNR among original, virtual averaged, and Gaussian smoothed nonframe-averaged Cirrus data were compared. 95% confidence intervals are presented in the parentheses.

For objective assessment using image quality metrics, the mean SNR and CNR were statistically significantly improved after virtual averaging (SNR: 30.5 vs. 47.6 dB, CNR: 4.4 vs. 6.4 dB, original versus averaged, $P < 0.0001$ for both comparisons; Table 2). No statistically significant differences in SNR were detected between original and Gaussian smoothed images but Gaussian smoothed images showed significantly higher CNR compared with the original images (SNR: 30.5 vs. 33.0 dB, $P = 0.11$, CNR: 4.4 vs. 6.7 dB, $P < 0.0001$, original versus Gaussian smoothed). Comparing the effects between virtual averaging and Gaussian smoothing, virtual averaged images presented statistically significantly higher SNR but significantly lower CNR comparing with Gaussian smoothed images (SNR: 33.0 vs. 47.6 dB, $P < 0.0001$, CNR: 6.4 vs. 6.7 dB, $P = 0.0015$, averaged versus Gaussian smoothed).

The distance between the end of visible nasal RNFL and the foveola (dNFL) was statistically significantly different between Cirrus and Spectralis before processing (681.4 vs. 446.5 μm, original Cirrus versus Spectralis, $P < 0.0001$) but after virtual averaging, no significant differences in dNFL between Cirrus and Spectralis was observed (442.9 vs. 446.5 μm, averaged Cirrus versus Spectralis, $P = 0.76$). No statistically significant difference in dNFL was observed between original Cirrus and Gaussian smoothed images (681.4 vs. 685.7 μm, original Cirrus versus Gaussian smoothed, $P = 0.78$). However, Gaussian smoothed images showed significantly longer dNFL comparing with Spectralis and virtual averaged Cirrus images (685.7 vs. 446.5 and 442.9 μm, Gaussian smoothed versus Spectralis and averaged Cirrus, both $P < 0.0001$). For the intraobserver repeatability assessment, the CVs of dNFL measurements were 3.6% for original Cirrus, 5.4% for original Spectralis, and 4.7% for averaged Cirrus data. The

CVs for the interobserver reproducibility were 9.2%, 13.5%, and 11.9% for original Cirrus, Spectralis, and averaged Cirrus data, respectively.

Experiment II: Effects of Signal Normalization and Minimizing Tissue Thickness Measurement Differences

For sectoral macular total retinal thicknesses, significant systematic measurement differences were detected in all sectors between Cirrus and Spectralis on both device outputs (Comparison I, Table 3, $P < 0.0001$) and between our self-designed software measurements before normalization (Comparison II and III, Table 3, $P < 0.0048$). After signal normalization, no significant differences were found in any of the sectors between Cirrus and Spectralis data except for in the outer temporal, outer nasal, and inner inferior sectors (Comparison IV, Table 3). Signal normalization significantly reduced the absolute differences between the devices in all sectors except for the center (mean absolute difference 20.3 μm [devices] to 6.7 μm [normalized], $P < 0.0001$).

For the cpRNFL thickness, statistically significant differences were found between Cirrus and Spectralis in the original device outputs and when using our custom segmentation algorithm before signal normalization (Comparison I, II, and III, Table 4). Although the differences found in the device outputs were significant, the detected differences (3.6 μm) were within the expected device measurement variability (4–5 μm) and were clinically nonsignificant (Comparison I).[15,36,37] Applying the same segmentation software (i.e., our custom segmentation software) made the differences larger (Comparison II and III, Table 4). After signal normalization, no systematic measurement differences were detected between Cirrus and Spectralis (Comparison IV, Table 4). Although the mean absolute difference between Cirrus and Spectralis was larger than the difference between device outputs (5.5 vs. 3.6 μm, Comparison IV versus I), the mean absolute differences after signal normalization were still subclinical and within the expected measurement variability, indicating that the proposed signal normalization did not add any significant artifacts.

Discussion

The proposed virtual averaging method significantly improved the image quality of nontracking nonframe-averaged OCT images and made the images comparable to active eye-tracking frame-averaged OCT images. By resampling voxel within a 3×3 neighborhood, adding a Gaussian deviation in voxel value multiple times, and then calculating the average, the proposed method successfully mimicked the way active eye-tracking frame-averaged devices acquiring images, and further reduced the measurement differences between nonframe-averaged and frame-averaged OCT images.

In virtual averaging, we assumed the deviation of the OCT signal during image acquisition came from the slight relocation of the camera, eye movement, as well as the variation in signal intensity from the same location caused by other factors, such as corneal dryness (e.g., before and after blinking) and slight differences in the incident angles; and the deviation of OCT signal was randomly distributed following Gaussian distributions. By adding Gaussian deviations in both sampling location and signal intensity, and repeating the process (Steps 1 and 2 in Fig. 1) multiple times, the outcome of virtual averaging showed much less noise, and thus improved the image quality and intraretinal layer contrasts. Retinal tissues like ELM and the end of nasal side RNFL, which are usually difficult to identify with confidence in the nonframe-averaged images, became clearly visible and easier to delineate after virtual averaging. By enhancing visualization of nonframe-averaged OCT images, the proposed method may help detect the fine structural changes in those originally obscure tomographic features and improve interpretation and assessment of the progression of pathologies; furthermore, it may enable detailed retinal structure studies on images, which previously fell short because of image quality, and may also enable more robust and finer retinal tissue segmentation.

Another potential benefit of the virtual averaging method is that it allows generating images similar to the ones with frame averaging with less image acquisition time. One practical drawback of eye-tracking system is the prolonged scanning time. The fast eye movement and blinking during scanning force the machine to repeat scanning at the same location until there is no detectable motion or blinking for the particular frame. The extended scanning duration makes certain scan patterns impractical especially for the densely sampled three-dimensional scans. With the proposed virtual averaging method, frame-averaged-like images can be generated from nonframe-averaged image data without prolonging the image acquisition time, which is especially beneficial to elderly and pathologic eyes.

Table 3. Sectoral Macular Total Retinal Thickness Measurements and Systematic Measurement Differences Between Cirrus and Spectralis, Using Four Comparison Methods

	Cirrus (μm)	Spectralis (μm)	Mean Absolute Differences (μm)	*P* Value
Comparison I (device outputs)				
Outer				
Temporal	258.3 (253.0, 263.6)	281.7 (276.4, 286.9)	23.3 (20.6, 26.1)	<0.0001*
Superior	278.0 (272.3, 283.6)	298.9 (292.7, 305.1)	21.0 (18.7, 23.2)	<0.0001*
Nasal	297.8 (290.2, 305.4)	315.9 (308.1, 323.7)	18.1 (15.3, 21.0)	<0.0001*
Inferior	266.6 (260.8, 272.4)	287.4 (281.0, 293.8)	20.8 (18.3, 23.3)	<0.0001*
Inner				
Temporal	308.9 (302.4, 315.3)	332.4 (327.3, 337.5)	23.5 (20.6, 26.5)	<0.0001*
Superior	322.4 (316.3, 328.5)	345.0 (339.4, 350.7)	22.7 (19.8, 25.5)	<0.0001*
Nasal	324.8 (318.9, 330.6)	346.9 (341.5, 352.3)	22.1 (19.1, 25.2)	<0.0001*
Inferior	317.9 (311.6, 324.1)	341.6 (336.2, 346.9)	23.7 (21.0, 26.5)	<0.0001*
Center	259.0 (252.7, 265.2)	275.0 (269.0, 280.9)	16.0 (13.5, 18.5)	<0.0001*
Comparison II (our custom algorithm with same parameters)				
Outer				
Temporal	288.4 (282.5, 294.4)	285.7 (280.8, 290.5)	4.5 (2.9, 6.1)	0.023*
Superior	304.5 (298.6, 310.5)	302.5 (297.6, 307.5)	5.3 (3.4, 7.2)	0.18
Nasal	321.4 (314.1, 328.7)	314.7 (308.3, 321.0)	7.7 (5.0, 10.3)	0.0002*
Inferior	293.4 (287.3, 299.5)	291.4 (285.1, 297.7)	6.8 (4.0, 9.6)	0.33
Inner				
Temporal	324.4 (319.0, 329.9)	322.0 (317.2, 326.8)	6.3 (4.2, 8.4)	0.16
Superior	343.0 (337.6, 348.4)	339.6 (334.5, 344.7)	6.0 (3.9, 8.0)	0.031*
Nasal	342.4 (337.5, 347.2)	337.5 (332.4, 342.6)	6.7 (4.7, 8.6)	0.002*
Inferior	338.1 (332.6, 343.6)	331.4 (326.5, 336.3)	7.7 (5.6, 9.8)	<0.0001*
Center	250.6 (243.0, 258.2)	261.7 (238.5, 285.0)	31.7 (12.5, 50.8)	0.34
Comparison III (our custom algorithm with fine tuned parameters)				
Outer				
Temporal	288.4 (282.5, 294.4)	280.1 (274.7, 285.4)	8.7 (6.4, 10.8)	<0.0001*
Superior	304.5 (298.6, 310.5)	294.3 (288.6, 300.0)	10.3 (8.2, 12.3)	<0.0001*
Nasal	321.4 (314.1, 328.7)	306.8 (300.0, 313.6)	15.1 (12.7, 17.5)	<0.0001*
Inferior	293.4 (287.3, 299.5)	283.1 (277.6, 288.6)	10.3 (8.0, 12.5)	<0.0001*
Inner				
Temporal	324.4 (319.0, 329.9)	318.0 (311.4, 324.7)	8.1 (4.6, 11.6)	0.0048*
Superior	343.0 (337.6, 348.4)	334.4 (329.1, 339.8)	9.1 (6.9, 11.3)	<0.0001*
Nasal	342.4 (337.5, 347.2)	333.5 (328.1, 339.0)	9.2 (7.2, 11.2)	<0.0001*
Inferior	338.1 (332.6, 343.6)	326.9 (321.6, 332.2)	11.2 (8.5, 14.0)	<0.0001*
Center	250.6 (243.0, 258.2)	236.6 (223.7, 249.6)	18.1 (8.5, 27.6)	0.0015*
Comparison IV (after signal normalization)				
Outer				
Temporal	291.4 (285.4, 297.3)	294.6 (289.5, 299.8)	5.0 (3.4, 6.7)	0.01*
Superior	307.0 (301.2, 312.9)	305.9 (300.5, 311.3)	3.8 (2.5, 5.1)	0.26
Nasal	325.4 (317.9, 332.9)	319.4 (313.1, 325.8)	7.2 (4.7, 9.8)	0.0009*
Inferior	296.7 (290.7, 302.8)	294.8 (289.4, 300.1)	4.9 (3.5, 6.3)	0.12
Inner				
Temporal	328.0 (321.7, 334.3)	328.2 (322.2, 334.2)	6.6 (4.1, 9.1)	0.93
Superior	346.4 (341.1, 351.7)	344.3 (339.2, 349.4)	4.5 (2.9, 6.0)	0.08

Table 3. Continued.

	Cirrus (μm)	Spectralis (μm)	Mean Absolute Differences (μm)	*P* Value
Nasal	345.7 (340.0, 351.3)	344.5 (339.0, 349.9)	4.5 (3.3, 5.6)	0.29
Inferior	341.2 (335.5, 346.9)	337.9 (332.9, 342.9)	4.7 (2.4, 7.0)	0.02*
Center	249.7 (240.1, 259.2)	251.1 (238.0, 264.2)	13.1 (8.2, 18.1)	0.71

95% CI are shown in the parentheses.

* Statistically significant differences were detected between Cirrus and Spectralis.

Although it is assumed that the deviation of the OCT signals in frame-averaged images was Gaussian distributed, applying Gaussian smoothing with the same kernel size to nonframe-averaged images could not reach the same effects as virtual averaging. Although the conventional Gaussian smoothing is very similar to the proposed virtual averaging, it applies smoothing evenly across the neighboring points. On the other hand, virtual averaging does not guarantee the all the neighboring information is used because of its random nature and the limited number of processing repetition. Gaussian smoothing reduced image noise effectively but smeared image details at the same time. Therefore, Gaussian smoothed images showed significant improvement in the CNR (measuring the smoothness of the image) but not in the SNR compared with the original images. Moreover, there were still significant differences in dNFL between Gaussian smoothed and Spectralis images. Those results imply that simply applying Gaussian smoothing is insufficient to resolve the discrepancies between nonframe-averaged and frame-averaged images. It is further supported by the results that modeling deviations of sampling location and amplitude and repeating the process are the keys to making nonframe-averaged OCT images comparable to frame-averaged OCT images by virtual averaging.

A potential limitation of the proposed virtual averaging is that the random resampling in the first step was in the x (transverse) and y (vertical) direction, but not in the z (axial) direction. By using the z-corrected nonframe-averaged Cirrus images, the deviation in the z direction was assumed to be negligible, and therefore, we randomly selected a voxel in a 2D 3 $\times$ 3 neighborhood in x and y directions instead of a three-dimensional 3 $\times$ 3 $\times$ 3 neighborhood. Although no big eye motion in the z direction were observed in the current dataset and no blurry effects in the retinal layer boundaries were detected by subjectively evaluating the virtual averaged images, the proposed method may not be suitable for the image with big z motion. In that case, aligning adjacent B-scans by using the ILM or RPE as the reference point would be needed before the virtual averaging processing. Further investigation is warranted.

Another potential limitation is that the output of the virtual averaging differs every time it runs due to its random resampling. However, we did not detect any noticeable difference in both qualitative and quantitative assessments on the multiple results of the virtual averaging on the same source image. Therefore, it is safe to assume that the random nature of the virtual averaging does not have adverse effect the outcome.

In addition to conventional image quality metrics, we used a new image quality metric, dNFL, to assess the OCT image quality before and after virtual averaging. It was observed that the visibility of the end of the nasal RNFL in fovea is sensitive to the image quality, and more specifically to the layer border

Table 4. Global Mean Circumpapillary RNFL Thickness Measurements and Systematic Measurement Differences Between Cirrus and Spectralis, Using Four Comparison Methods

Methods	Cirrus (μm)	Spectralis (μm)	Mean Absolute Differences (μm)	*P* Value
Comparison I	96.3 (91.2, 101.3)	99.1 (94.8, 103.4)	3.6 (2.3, 5.0)	0.003*
Comparison II	99.4 (94.5, 104.3)	111.3 (107.2, 115.3)	11.9 (9.3, 14.4)	<0.0001*
Comparison III	99.4 (94.5, 104.3)	106.1 (102.2, 110.1)	7.2 (5.2, 9.3)	<0.0001*
Comparison IV	101.9 (97.0, 106.7)	100.4 (96.6, 104.2)	5.5 (3.7, 7.2)	0.20

95% CI are shown in the parentheses.

* Statistically significant differences were detected between Cirrus and Spectralis.

clarity and contrast. Even though dNFL measurements quantified the image quality from a more perceptual perspective and its subjective nature may limit the legitimacy, it showed good intraobserver repeatability (CVs range 3.6%–5.4%) and interobserver reproducibility (CVs range 9.2%–13.5%), indicating that it could be considered as a reliable image quality metric. Spectralis data had the largest CVs in both intraobserver repeatability and interobserver reproducibility, which may be caused by variations in OCT signal characteristics such as hyperreflection signals. Despite its subjective nature, dNFL measurements are practical indicators of the visibility and clarity of intraretinal layers, especially for the RNFL. Therefore, dNFL measurements assess more clinical aspect on the performance of virtual averaging method.

Besides being able to enhance image quality, the virtual averaging can be used to normalize OCT signal between nonframe-averaged and frame-averaged images so that the systematic differences in quantitative measurements can be minimized. Because different segmentation algorithms were used to measure the retinal thicknesses and the foveola was not always located at the center of the scan window, there was a speculation that the systematic measurement differences in retinal thicknesses between OCT devices could be minimized by applying identical segmentation algorithm and adjusting the foveola position. However, our results showed that even using the same segmentation algorithm and aligning the foveola position, significant differences were detected in the sectoral macular total retinal thickness between Cirrus and Spectralis (Comparison II and III, Table 2). Although the absolute difference was reduced from the differences with the original device measurements, it indicated that the signal normalization process was still required. After applying signal normalization and using our self-designed segmentation algorithm, the systematic measurement differences between nonframe-averaged and frame-averaged OCT data were successfully reduced and the clinical measurements of sectoral macular total retinal thicknesses from them were made directly comparable.

The cpRNFL thickness showed a significant difference in the original device outputs between Cirrus and Spectralis. Though the absolute difference did not reach clinical significance as reported in the literature,[15,36,37] there was a consistent and statistically significant trend of Spectralis measurements being thicker than the Cirrus measurements. Similar to the results found between nonframe-averaged OCT devices,[19] the systematic measurement differences in cpRNFL thickness were not reduced by applying the same segmentation algorithm, but became even larger, indicating that signal normalization is needed to minimize the measurement differences. Despite cpRNFL thickness not needing any post hoc processing to make it clinically comparable between Cirrus and Spectralis,[11] the statistically significant trend between Cirrus and Spectralis disappeared after signal normalization, and the mean absolute differences were still subclinical and within the measurement variability range, suggesting that the signal normalization process did not add any significant adverse noise or artifacts.

In conclusion, the novel virtual averaging method can be a fundamental image processing technique that enhances image quality without the need of hardware eye-tracking system, bridges the gap between nonframe-averaged and frame-averaged images, and makes both qualitative and quantitative assessments between nonframe-averaged and frame-averaged OCT images directly comparable. This method may enable detailed retinal structure studies on images acquired using a mixture of nonframe-averaged and frame-averaged OCT devices without concerning about systematic differences in both qualitative and quantitative aspects.

Acknowledgments

Supported in part by National Institutes of Health Grants NIH R01-EY013178, R01-EY011289, P30 EY008098 (Bethesda, MD), the Eye and Ear Foundation (Pittsburgh, PA), and Research to Prevent Blindness (New York, NY).

Presented in part at the annual meeting of the Association for Research in Vision and Ophthalmology, Seattle, Washington, May 2013.

C-LC and HI contributed equally to the work and therefore should be considered equivalent authors.

Disclosure: **C.-L. Chen,** None; **H. Ishikawa,** None; **G. Wollstein,** None; **R.A. Bilonick,** None; **L. Kagemann,** None; **J.S. Schuman,** Zeiss (P)

References

1. Huang D, Swanson E, Lin C, et al. Optical coherence tomography. *Science*. 1991;254:1178–1181.

2. Drexler W, Fujimoto JG. State-of-the-art retinal optical coherence tomography. *Prog Retin Eye Res*. 2008;27:45–88.
3. Schuman JS, Puliafito CA, Fujimoto JG, Duker JS. Optical Coherence Tomography of Ocular Diseases. Thorofare, NJ: *SLACK Incorporated*; 2013.
4. Marschall S, Sander B, Mogensen M, Jorgensen TM, Andersen PE. Optical coherence tomography-current technology and applications in clinical and biomedical research. *Anal Bioanal Chem*. 2011;400:2699–2720.
5. Wojtkowski M, Leitgeb R, Kowalczyk A, Bajraszewski T, Fercher AF. In vivo human retinal imaging by Fourier domain optical coherence tomography. *J Biomed Opt*. 2002;7:457–463.
6. Baumal CR. Clinical applications of optical coherence tomography. *Curr Opin Ophthalmol*. 1999;10:182–188.
7. Costa RA, Skaf M, Melo LA Jr, et al. Retinal assessment using optical coherence tomography. *Prog Retin Eye Res*. 2006;25:325–353.
8. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol*. 1995; 113:586–596.
9. Buchser NM, Wollstein G, Ishikawa H, et al. Comparison of retinal nerve fiber layer thickness measurement bias and imprecision across three spectral-domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2012;53:3742–3747.
10. Giani A, Cigada M, Choudhry N, et al. Reproducibility of retinal thickness measurements on normal and pathologic eyes by different optical coherence tomography instruments. *Am J Ophthalmol*. 2010;150:815–824.
11. Leite MT, Rao HL, Weinreb RN, et al. Agreement among spectral-domain optical coherence tomography instruments for assessing retinal nerve fiber layer thickness. *Am J Ophthalmol*. 2011;151:85–92.e1.
12. Kanamori A, Nakamura M, Tomioka M, Kawaka Y, Yamada Y, Negi A. Agreement among three types of spectral-domain optical coherent tomography instruments in measuring parapapillary retinal nerve fibre layer thickness. *Br J Ophthalmol*. 2012;96:832–837.
13. Bourne RR, Medeiros FA, Bowd C, Jahanbakhsh K, Zangwill LM, Weinreb RN. Comparability of retinal nerve fiber layer thickness measurements of optical coherence tomography instruments. *Invest Ophthalmol Vis Sci*. 2005;46: 1280–1285.
14. Monteiro ML, Leal BC, Moura FC, Vessani RM, Medeiros FA. Comparison of retinal nerve fibre layer measurements using optical coherence tomography versions 1 and 3 in eyes with band atrophy of the optic nerve and normal controls. *Eye (Lond)*. 2007;21:16–22.
15. Vizzeri G, Weinreb RN, Gonzalez-Garcia AO, et al. Agreement between spectral-domain and time-domain OCT for measuring RNFL thickness. *Br J Ophthalmol*. 2009;93:775–781.
16. Sung KR, Kim DY, Park SB, Kook MS. Comparison of retinal nerve fiber layer thickness measured by Cirrus HD and Stratus optical coherence tomography. *Ophthalmology*. 2009; 116:1264–1270. 70.e1.
17. Knight OJ, Chang RT, Feuer WJ, Budenz DL. Comparison of retinal nerve fiber layer measurements using time domain and spectral domain optical coherent tomography. *Ophthalmology*. 2009;116:1271–1277.
18. Chen CL, Ishikawa H, Wollstein G, et al. Individual A-scan signal normalization between two spectral domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2013;54: 3463–3471.
19. Chen CL, Ishikawa H, Ling Y, et al. Signal normalization reduces systematic measurement differences between spectral-domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2013;54:7317–7322.
20. Martinez-Conde S, Macknik SL, Hubel DH. The role of fixational eye movements in visual perception. *Nat Rev Neurosci*. 2004;5:229–240.
21. Klein T, Wieser W, Eigenwillig CM, Biedermann BR, Huber R. Megahertz OCT for ultrawide-field retinal imaging with a 1050 nm Fourier domain mode-locked laser. *Opt Express*. 2011;19:3044–3062.
22. Potsaid B, Gorczynska I, Srinivasan VJ, et al. Ultrahigh speed spectral / Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second. *Opt Express*. 2008;16:15149–15169.
23. Pemp B, Kardon RH, Kircher K, Pernicka E, Schmidt-Erfurth U, Reitner A. Effectiveness of averaging strategies to reduce variance in retinal nerve fibre layer thickness measurements using spectral-domain optical coherence tomography. *Graefes Arch Clin Exp Ophthalmol*. 2013;251: 1841–1848.
24. Hangai M, Yamamoto M, Sakamoto A, Yoshimura N. Ultrahigh-resolution versus speckle

noise-reduction in spectral-domain optical coherence tomography. *Opt Express*. 2009;17:4221–4235.
25. González RC, Woods RE. Digital image processing. Upper Saddle River, NJ: Prentice Hall; 2002.
26. Chitchian S, Fiddy M, Fried NM. Wavelet denoising during optical coherence tomography of the prostate nerves using the complex wavelet transform. *Conf Proc IEEE Eng Med Biol Soc*. 2008;2008:3016–3019
27. Puvanathasan P, Bizheva K. Interval type-II fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images. *Opt Express*. 2009;17:733–746.
28. Puvanathasan P, Bizheva K. Speckle noise reduction algorithm for optical coherence tomography based on interval type II fuzzy set. *Opt Express*. 2007;15:15747–15758.
29. Adler DC, Ko TH, Fujimoto JG. Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter. *Opt Lett*. 2004;29:2878–2880.
30. Ozcan A, Bilenca A, Desjardins AE, Bouma BE, Tearney GJ. Speckle reduction in optical coherence tomography images using digital filtering. *JOSA A*. 2007;24:1901–1910.
31. Ferris FL III, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol*. 1982;94:91–96.
32. Ishikawa H, Stein DM, Wollstein G, Beaton S, Fujimoto JG, Schuman JS. Macular segmentation with optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2005;46:2012–2017.
33. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. 2012.
34. Boker S, Neale M, Maes H, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika*. 2011;76:306–317.
35. Bilonick R. merror: Accuracy and Precision of Measurements. R package version 1.0; 2003.
36. Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology*. 2009;116:1257–1263. 63.e1-2.
37. Tan BB, Natividad M, Chua KC, Yip LW. Comparison of retinal nerve fiber layer measurement between 2 spectral domain OCT instruments. *J Glaucoma*. 2012;21:266–273.