SIGNAL NORMALIZATION AMONG MULTIPLE OPTICAL COHERENCE TOMOGRAPHY DEVICES

by

Chieh-Li Chen

Bachelor of Science in Electrical Engineering, National Taiwan University, 2008

Master of Science in Electronics Engineering, National Taiwan University, 2010

Submitted to the Graduate Faculty of

The Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Bioengineering

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Chieh-Li Chen

It was defended on

September 8th, 2014

and approved by

Howard Jay Aizenstein, MD, PhD, Associate Professor, Department of Psychiatry

Larry Kagemann, PhD, Associate Professor, Department of Bioengineering

Kang Kim, PhD, Associate Professor, Departments of Medicine and Bioengineering

Joel S. Schuman, MD, Distinguished Professor, Department of Ophthalmology

George Stetten, MD, PhD, Professor, Department of Bioengineering

Gadi Wollstein, MD, Associate Professor, Department of Ophthalmology

Dissertation Director: Hiroshi Ishikawa, MD, Assistant Professor, Departments of

Ophthalmology and Bioengineering

SIGNAL NORMALIZATION AMONG MULTIPLE OPTICAL COHERENCE TOMOGRAPHY DEVICES

Chieh-Li Chen, PhD

University of Pittsburgh, 2014

Optical coherence tomography (OCT) has become a clinical standard in ophthalmology because it has the ability to provide *in vivo* cross-sectional images of ocular tissues with microscopic resolution in a non-contact and non-invasive manner. More and more manufacturers are getting involved in the race of instrument design and the development of the spectral-domain OCT (SD-OCT). Various light sources, optical designs, and image acquisition settings were employed by different manufacturers to stand out among competitors. This provides a wide variety of options in terms of scanning protocol, image processing, and presentation. However, the diversity also reflects in the variability in the OCT signal characteristics. The variability of OCT signal characteristics not only results in systematic differences in OCT measurement data, such as the retinal nerve fiber layer (RNFL) thickness and total retinal thickness, but also induces discrepancies in OCT image appearance. Those differences cause serious clinical challenges when comparing OCT images from different OCT devices, or recruiting multiple OCT devices in one study.

To solve this problem, a novel signal normalization method was developed in this dissertation. The signal normalization was developed in a stepwise fashion to resolve all factors contributing to the systematic differences among various OCT devices, including axial sampling density, the amount of speckle noise, intensity dynamic range, and image quality. Quantitative analyses and qualitative assessments were conducted to evaluate the proposed signal

normalization method. For the quantitative analyses, engineering and clinical validations were performed via measuring the absolute differences in A-scan profile intensity and comparing the systematic RNFL thickness differences before and after signal normalization. For the qualitative assessment, subjective evaluation of the similarity of OCT image appearance through a questionnaire was performed. Statistically significant reduction in both the absolute difference in A-scan profile and the systematic differences among SD-OCT devices were observed after signal normalization. Statistically significant improvements of image similarity between OCT image pairs were also found after the processing. With the proposed signal normalization method, quantitative analysis as well as qualitative assessment among OCT devices will become directly comparable, which would broaden the use of OCT technology in both clinical and research applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

OCT – Optical coherence tomography

TD-OCT – Time-domain optical coherence tomography

SD-OCT – Spectral-domain optical coherence tomography

A-scan – Axial scan

FD-OCT – Fourier-domain optical coherence tomography

SS-OCT – Swept-source optical coherence tomography

CCD – Charge-coupled device

3D – Three-dimensional

MRI – Magnetic resonance imaging

CT – Computed tomography

UBM – Ultrasound biomiscoscopy

AMD – Age-related macular degeneration

ILM – Internal limiting membrane

RPE – Retinal pigment epithelium

RNFL – Retinal nerve fiber layer

ONH – Optic nerve head

AS-OCT – Anterior segment optical coherence tomography

LASIK – Laser in situ keratomileusis

DICOM – Digital Imaging and Communication in Medicine

SEM – Structural equation model

SS – Signal strength

SSI – Signal strength index

CSLO – Confocal laser scanning ophthalmoscope

HDR – High-dynamic range

2D – Two-dimensional

SNR – Signal-to-noise ratio

CNR – Contrast-to-noise ratio

GC-IPL – Ganglion cell and inner plexiform layer

IN-OPL – Inner nuclear and outer plexiform layer

SD – Standard deviation

CI – Confidence interval

HM – Histogram matching

PET – Positron emission tomography

SPECT – Single photon emission computed tomography

OPL – Outer plexiform layer

ELM – External limiting membrane

GCL – Ganglion cell layer

IS/OS – Inner and outer segment

CoV – Coefficient of variation

IPL – Inner plexiform layer

QI – Quality index

MD – Mean deviation

dNFL – Distance between the end of visible nasal RNFL and the foveola

ETDRS – Early Treatment Diabetic Retinopathy Study

ONL – Outer nuclear layer

PDR – Proliferative diabetic retinopathy

NPDR – Non-proliferative diabetic retinopathy

DME – Diabetic macular edema

CME – Cystoid macular edema

FDA – U.S. Food and Drug Administration

PREFACE

Four years have passed by and I still remember the day when I got my admission letter from the University of Pittsburgh and the first day when I joined the Ophthalmic Imaging Research Laboratory, also known as Glaucoma Imaging Group (GIG). Those days are so far away but they also felt like yesterday. I still remember how excited and nervous I was about the new journey set in front of me. I am so lucky to join the most talented and amazing group to explore the small but beautiful world of the eye.

Thank you, Dr. Ishikawa, for all of your guidance and support. Ian once told me that I have an excellent mentor, and I simply cannot agree with him more. The luckiest thing in my PhD study was to have you as my mentor and as a friend. I could never have appreciated how fun doing research is and could not have written this acknowledgement without your patient mentoring. I will keep improving myself toward the goal of becoming an independent researcher!

Thank you, Dr. Schuman, for providing us a wonderful research environment to test all our crazy hypotheses, and for inspiring many more audacious ideas. Thank you, Dr. Wollstein, for leading this amazing group and showing us how rigorous we should be when conducting any experiment. Thank you, Dr. Bilonick, for all your help on the statistical analyses, and the patience you have shown me when I kept asking silly statistical questions. Thank you, Dr. Kagemann, for sharing with us the latest news and results in ophthalmology research. Thank

1.0   INTRODUCTION

Optical coherence tomography (OCT) is a non-invasive and non-contact optical imaging technique. It generates *in-vivo* cross-sectional images of biological structures with microscopic resolution in a real-time fashion by measuring the interference of the reflected signals from the reference mirror and from the tissue.[1, 2] Early OCT systems detect the correlation of the light echoes in time domain using interferometry, and therefore are referred to as time-domain OCT (TD-OCT).[1-3] Instead of the physical reference mirror, the later OCT generation detects the backscattering signals in frequency domain with a spectrometer, encodes the time delay information as a spectrum, and therefore are often referred to as spectral-domain OCT (SD-OCT).[3, 4] The hardware improvement in SD-OCT furthermore enhances the axial resolution, improves the image acquisition speed, and allows volumetric image acquisition.[2, 3, 5] With the advances of OCT technology in both hardware design and software developments, OCT has become an indispensable tool in clinical routines for both the qualitative and quantitative assessment of tissue structures to help disease diagnoses and management, especially in ophthalmology.[6]

Multiple SD-OCT devices have been made commercially available from several different manufacturers. Each device is equipped with different optical design and image acquisition settings, providing a wide variety of options in terms of scanning protocol, image processing, and presentation. This diversity, however, not only results in OCT image data incompatibility,

but also generates substantial differences in both qualitative and quantitative OCT image measurements/interpretations among the various OCT devices. The measurement differences pose a serious clinical challenge, as OCT measurements obtained using different OCT devices or different device generations are not directly comparable. In addition, the discrepancies in the appearance of biological structures require extra caution be taken when qualitatively evaluating the OCT images. These factors limit the uses of OCT devices because they both lead to inflexibility with regard to switching between various OCT devices or generations when patients move from one clinic to another or device models update or change.

To solve this problem, a signal normalization method was proposed and developed in this dissertation. The ultimate goal is to develop a signal normalization method to minimize the discrepancies in signal characteristics among multiple OCT devices, and to make the outcomes (both OCT image appearance and measurements) directly comparable so that clinicians can fully utilize OCT technology without worrying about which device was used for image acquisition and compare measurements and appearance obtained from different devices directly.

1.1 OPTICAL COHERENCE TOMOGRAPHY

The conventional OCT technique typically consists of an interferometer with a low coherence and a broad bandwidth light source (Figure 1). The light is split into two beams after passing through a beam splitter. One beam is sent to the reference arm which has a reference mirror, and the other beam is sent into the biological samples. The combination of the backscatter signal from the samples and the reflective signal from the reference mirror generates an interference pattern with maximal intensity if the path length to the reference mirror and tissue are within the

coherence length of the light source.[3, 7, 8] Depth-resolved tissue reflectivity characteristics are then extracted from the interference pattern and recorded as an intensity profile in the axial direction. By changing the position of the reference mirror, backscattered tissue intensity can be detected from different depths in the tissue sample. A sequence of echoes from a single retina location are presented in a series as an axial scan (A-scan), and optical cross-sections (B-scans) are obtained by scanning the OCT beam in the transverse direction (Figure 2). Since time-encoded signals are obtained in this manner, this approach is referred to as TD-OCT.[3]
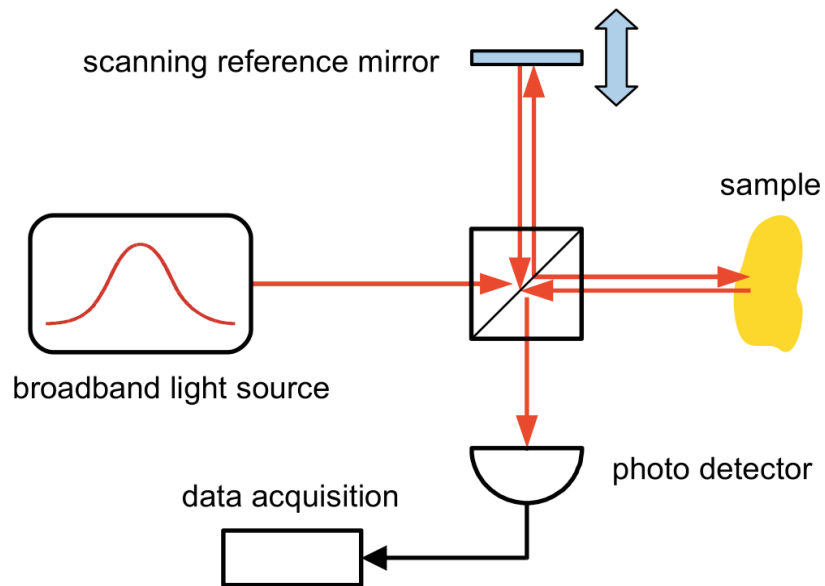
Figure 1. The schematic figure of TD-OCT instrument. (Marschall et al. [9])
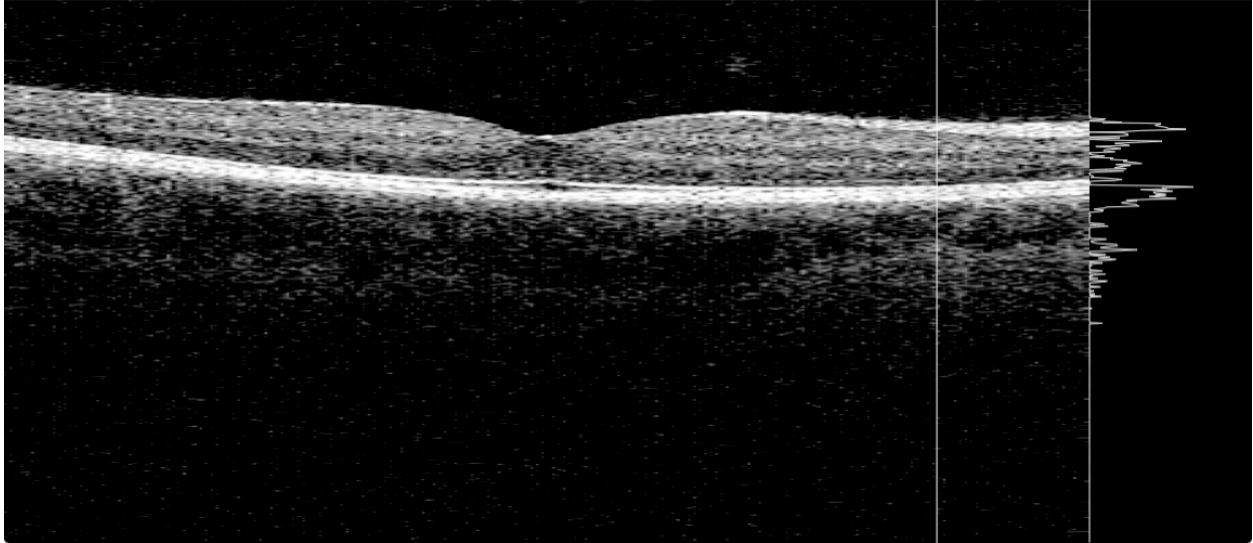
Figure 2. An example of cross-sectional image (B-scan) through a healthy macula. The vertical white line indicates the location of the single A-scan intensity profile shown on the right.

Fourier-domain OCT (FD-OCT) includes SD-OCT and swept-source OCT (SS-OCT). Instead of using a physically moving reference mirror, FD-OCT detects the backscattering signals from the biological tissues in the frequency domain with either a combination of a broad-bandwidth light source, charge-coupled device (CCD) camera, and a spectrometer (SD-OCT), or by sweeping through a range of frequencies (SS-OCT) (Figure 3).[10-16] Frequency information from all depths at a given point (about 20 μm in diameter) in the tissue is acquired simultaneously and converted into an intensity profile by Fourier-transform. The implementation of the broadband light source with broader bandwidth enhances the axial resolution from ~10 μm to 2 μm, and the introduction of the spectrometer or sweeping frequencies improves the image acquisition speed (from 400 A-scans/s to between 26,000 and 100,000 A-scans/s).[2, 5, 17, 18] In addition to faster scanning speed, a higher signal-to-noise ratio is offered with a perfect reflector.[14] With the improvements of scanning speed and resolution in SD-OCT and SS-OCT, collecting volumetric (three-dimensional; 3D) scans of biological tissues becomes feasible. More

information from the retina can be acquired in a relatively short amount of time, allowing more detailed biological structure visualization.[11, 19-22] Although SS-OCT is able to provide a faster scanning rate, less photons reaching the retinal tissues results in the potential limitation of a lower signal to noise ratio, and the choice of a tunable narrowband laser light source further reduces the resolution of SS-OCT.[3]
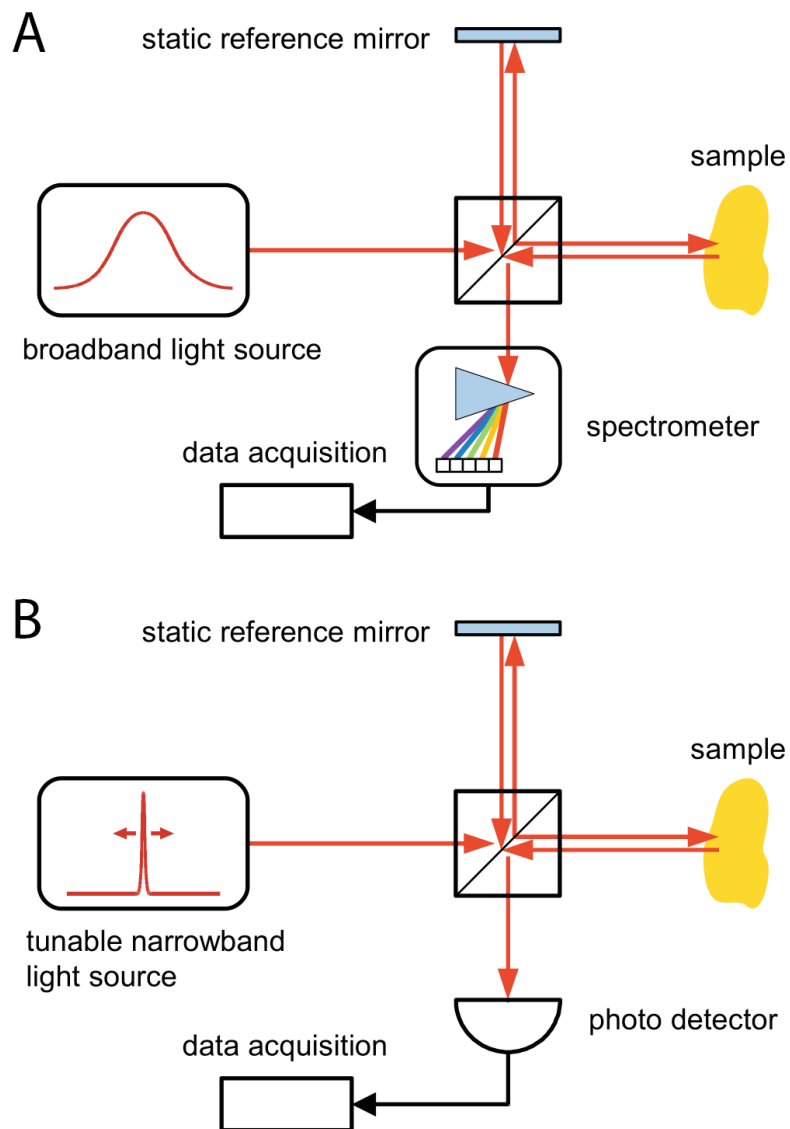
Figure 3. The schematic illustration of FD-OCT instrument. (A) SD-OCT and (B) SS-OCT. (Marschall et al. [9])

5

1.2   OPTICAL COHERENCE TOMOGRAPHY APPLICATIONS IN

OPHTHALMOLOGY

With the invention of OCT, medical diagnostic imaging technology has entered a new era. The resolution of OCT is superior to other *in vivo* imaging techniques such as magnetic resonance imaging (MRI), computed tomography (CT), scanning laser ophthalmoscopy, B-mode ultrasound, and ultrasound biomicroscopy (UBM, or high frequency ultrasound), which have image resolutions of 1 mm, 300, 300, 150, and 20 μm, respectively.[23-25] Its non-invasive and non-contact nature has made OCT imaging become known as "optical biopsy." With the ability to perform "optical biopsy" with microscopic resolution (as high as 2 μm), OCT improved not only the understanding of disease pathogenesis,[24-26] but also clinical diagnosis and management of various diseases.[18, 24-29]

The major medical applications of OCT focus on tissues that are accessible by optics and need high spatial resolution cross-sections, such as eye, skin, surgically exposed tissue, and surfaces that can be reached by endoscopic probes. Among those clinical applications, OCT is particularly suitable for ophthalmology because of the optical properties and transparent nature of the eye, and the accessibility of the retina to transpupillary examination.[24] In addition, opposed to fluorescein angiograph and ultrasound imaging, which may cause considerable discomfort to the patient, OCT is noninvasive and requires no physical contact with the eye. All these advantages have led to the rapid success of OCT in ophthalmology. Since its first application in ophthalmology, OCT has characterized a wide variety of retinal pathology and the images correspond with the histopathologic features of these disorders.[24, 30]

Retinal and macular diseases, such as macular holes, diabetic retinopathy, and age-related macular degeneration (AMD), used to be diagnosed with fundoscopic exam, fluorescein

angiography, and B-scan ultrasound. However, insufficient resolution prohibits the instruments from providing detailed information on retinal structures. The microscopic resolution of OCT imaging allows the discrimination of the retina's micron-scale layered structures; and the cross-sectional images are able to present corresponding histopathologic features of the diseases.[24, 30] OCT has proven to be very helpful in the evaluation of retinal pathologies, elucidating pathological processes, assessing if surgical intervention is required, or in monitoring the outcome of surgery. For example, OCT data help detect the primary pathogenic event in idiopathic macular hole formation, which appears as a localized perifoveal vitreous detachment;[31-34] OCT-derived measurements become useful information for assessing the anatomical and functional success after surgery.[26] The non-invasive characteristics of OCT, which produces less discomfort to the patients, has created a trend towards the less frequent use of fluorescein angiography, especially during frequent follow-up visits.[26] OCT technology also helps the evaluation of the vitreorretinal interface, such as the epiretinal membrane. An attached epiretinal membrane can be appreciated on OCT as a contrast in reflectivity between the membrane and retina, or as the presence of an edge, or as a steepened foveal contour.[26] The morphological information of the retinal tissues provided by OCT images further plays an important role in the disease diagnoses and managements for macular edema and AMD. Qualitative OCT assessment demonstrates subtle disease activity before biomicroscopy fundus examination and visual acuity changes; moreover, the ability to detect the presence of intraretinal fluid in OCT imaging shows an effective way to guide treatment and retreatment, as intraretinal fluid is associated with the presence of neovascular membranes in AMD.[35] Besides the qualitative assessment of OCT images, the quantitative measurements, such as the total retinal thickness (the thickness between internal limiting membrane (ILM) and retinal pigment

epithelium (RPE)) in various macular locations, and total macular volume further provide objective information for disease monitoring and for evaluating responses to therapy.[36]

Glaucoma, one of the leading causes of blindness worldwide, is a slowly progressing optic neuropathy that is characterized by the loss of retinal ganglion cells and the retinal nerve fiber layer (RNFL) with associated visual field loss.[37, 38] Because optic nerve damages are irreversible, it is crucial to detect glaucoma in its earliest possible stage and monitor disease progression closely so that proper interventions are applied at the right time to prevent further damage. The ability to provide an objective quantitative assessment of RNFL thickness has made OCT part of the clinical routine for glaucoma practice. Total retinal and circumpapillary RNFL thicknesses are measured via the automated segmentation software embedded in the commercialized OCT device. The measurements of optic nerve head (ONH) parameters, such as cup area, disc area, and cup-to-disc ratio were achieved later when 3D volume scans can also be obtained in both TD-OCT and FD-OCT images. Good reproducibility in measuring the RNFL thickness and ONH parameters has been observed.[39-42] Several previous studies also have demonstrated that OCT was capable of differentiating glaucomatous from healthy eyes.[27-29, 36, 43-47] Furthermore, OCT is able to capture both progressive RNFL thinning and ONH remodeling, which helps monitor glaucoma progression.[18, 48, 49]

Anterior segment OCT (AS-OCT) extends the use of the OCT imaging applications from the posterior to the anterior part of the eye by providing structural information about the cornea, anatomic structures of angle, structures not visible with slitlamp examination, and the anterior chamber.[50] Compared to the conventional UBM, AS-OCT enables the visualization and quantification of the anterior segment of the eye with high resolution in a non-invasive and non-contact manner in the clinic. Therefore, AS-OCT is frequently used not only for diagnosing

pathologic conditions of anterior segment, but also for evaluating anterior surgeries.[51-54] With AS-OCT, distinguishing the epithelial layer from the stroma and observing the wound-healing process of the cornea become possible.[50] Studies have shown that epithelial ingrowth under the laser in situ keratomileusis (LASIK) flap could be seen with AS-OCT. For quantitative assessments, objective measurement of cornea thickness shows reliable pachymetric mapping. Pachymetric measurements are useful for LASIK surgery and the diagnosis of keratoconus and other corneal pathologies.

In summary, OCT technology provides additional information of disease diagnosis and monitoring. OCT enables the subjective visualization and objective measurement of ocular tissues, including the cornea and retina, in a real-time, non-contact, and non-invasive fashion with microscopic axial resolution. With the advances in OCT technology and the improvement of computer-aided image processing tools, OCT has become an indispensable ocular imaging device in daily clinical care for ophthalmology. In this dissertation, we will focus on the retinal and RNFL thickness measurements and tissue structure visualization application of OCT, especially for the region around the ONH and macular region.

1.3   CLINICAL PROBLEMS TO BE SOLVED

Multiple SD-OCT devices from several manufacturers are commercially available currently (Table 1). While all manufacturers aim to improve OCT technology in order to improve their clinical diagnoses and disease detection abilities, each manufacturer also dedicates their efforts to advancing their products without concern towards data/measurement compatibility and

comparability. They use different optical designs, device settings, and image processing methods.[26, 55] This diversity causes OCT data inconsistency and OCT measurement incompatibility.

Table 1. List and technical characteristics of the commercially available SD-OCT.

Instrument	Manufacturer	Center wavelength (nm)	Axial / Transverse resolution (µm)	Scan speed (A-scans/s)	Data format (bits/pixel)	Scan depth (mm) / Number of samples
Cirrus HD-OCT	Zeiss	840	5 / 15	27,000	8	2.0 / 1024
RTVue OCT	Optovue	840	5 / 15	26,000	12	2-2.3 / 768
Spectralis OCT	Heidelberg Engineeriing	870	3.9 / 14	40,000	16	1.9 / 496

Significant differences in retinal layer thickness measurements were observed when comparing OCT measurements from different devices.[22, 56-59] The systematic measurement differences among OCT devices pose a serious challenge. RNFL thickness is an excellent biomarker and plays an important role for glaucoma diagnosis and monitoring.[18, 28, 36, 44, 47-49, 60] Because of its slowly progressing nature, a long-term follow-up is needed for early progression detection. During the follow-up period, if patients move from one clinic to another, or device models change or update, the systematic differences between various OCT devices or model iteration make previous data useless and prevent us from establishing a long-term clinical record of RNFL thickness measurements, which is required for glaucoma progression analysis. Besides, when it comes to multicenter clinical studies, OCT data coming from different clinical

centers may be acquired using different OCT devices or different generations. In that case, the data incompatibility among different OCT devices would make it difficult to compare the measurements directly, further reduce the amount of eligible data and limit OCT applications.

Inconsistent image appearance, such as contrast between retinal signal and background noise or contrast between adjacent intra-retinal layers, may mislead clinical diagnoses if one is evaluating images from different devices. In addition, the variations in OCT data format, such as sampling numbers and intensity dynamic range, make OCT image visualization machine-specific. Specific image browsers and parameter settings are needed when opening and reading OCT data from different devices. This adds complexity and inconvenience when browsing or comparing OCT images from different OCT devices. Although all devices can export OCT images in generic graphic formats (e.g. JPEG, TIFF, etc.), those data are usually down sampled, especially in sampling bit rate (12- or 16-bit data becomes 8-bit) together with possible compression artifacts, which further affects processing especially for detailed segmentation of intra-retinal layers. The Digital Imaging and Communication in Medicine (DICOM)[61] has put efforts into OCT image format standardization, but nevertheless, still cannot fully resolve the discrepancy among various devices. The data incompatibility and image discrepancy among multiple OCT devices cause a serious problem for both clinical practice and research study design. Therefore, an image processing method that is able to normalize the various signal characteristics and minimize the systematic differences among different SD-OCT devices would be indispensable.

1.4 SYSTEMATIC DIFFERENCES IN OPTICAL COHERENCE TOMOGRAPHY MEASUREMENTS

During the last 25 years, many models and generations have been developed and commercialized, such as OCT1, OCT 2000, and OCT3 (also known as Stratus (Carl Zeiss Meditec)) (all TD-OCTs), and the newly invented FD-OCT devices, which were developed by various manufacturers. Although new models offer faster scanning speed, improved scan quality, and advanced processing software, significant measurement differences are usually observed between the new models and the existing instruments, which poses a serious challenge when it comes to comparing the RNFL thickness measurements from various devices during a patients' follow-up periods.[56-58]

Several studies were conducted to investigate the compatibility of RNFL thickness measurements among the early generations of OCT, namely OCT1, OCT 2000, and Stratus. Bourne et al. found that the RNFL thickness measurements were thinner with Stratus compared with OCT 2000. Even after applying a correction factor, the variability exceeded TD-OCT's limit of resolution (10 μm).[62] Monteiro et al. compared the RNFL thickness measurements between OCT1 and Stratus, and found that the measurements were smaller with Stratus than with OCT1.[63] Both studies concluded that the measurement agreement among the TD-OCT devices were poor from a clinical standpoint and should not be used interchangeably.[62, 63]

Global, quadrant, and clock hour mean RNFL thickness measurements were compared among Stratus and various SD-OCT devices to test their compatibility as well. Systematic differences in RNFL thickness measurements were observed between Stratus and Cirrus (Zeiss, Dublin, CA, USA), between Stratus and RTVue (Optovue, Fremont, CA, USA), and between Stratus and Spectralis (Heidelberg Engineering, Heidelberg, Germany), respectively.[56, 64-66]

12

Vizzeri et al. found a proportional bias between Stratus and Cirrus, but in general Stratus consistently provided thicker RNFL measurements than Cirrus for both global and sectoral parameters even when no statistically significant difference in image quality was detected (image quality has been recognized as a main factor influencing thickness measurements).[64, 67-72] The same trend was observed by Knight et al. and Sung et al.[65, 66] Seibold et al. compared the RNFL thickness among Stratus, Cirrus, RTVue, and Spectralis, individually. They found that Stratus reported thicker RNFL thickness than Cirrus and Spectralis but thinner thickness than RTVue, and concluded that the measurement values were significantly different and clinicians should be aware of relationships between various OCT machines when following glaucoma patients, especially if switching instruments or comparing scans from various machines.[56] Giani et al. extended the measurements from the ONH to the macular region and compared the thickness measurements from Stratus with five SD-OCT devices (Cirrus, RTVue, Spectralis, SDOCT Coopernicus HR (Optopol Technology SA, Zawiercie, Poland), and 3D OCT (Topcon, Tokyo, Japan)).[57] They found that differences between OCT devices were beyond expected device measurement variability and should not be used interchangeably.[57] With the systematic differences, many pieces of literature suggested that standardized OCT measurements for clinical practice are needed and it is imperative that clinical trials use single OCT device for the disease follow-ups.[57]

Leite et al. tested the agreement of RNFL thickness assessment among SD-OCT instruments.[58] They found that RTVue and Spectralis reported statistically significant thicker RNFL thickness measurements as compared to Cirrus, and detected fixed biases and proportional biases in the measurement values. The 95% limit of agreement was around 20 μm, which was nearly twice the expected measurement variability within a single SD-OCT device (±5 μm).

13

They summarized that systematic differences existed among SD-OCT devices and suggested comparing with histologic measurements to help determine which technique was most accurate.[58]. Buchser et al. and Kanamori et al. compared the RNFL thickness among Cirrus, RTVue, and 3D OCT (Buchser et al. used 3D OCT-1000 while Kanamori et al. used 3D OCT-2000). In their findings, RTVue showed statistically significantly thicker RNFL measurements than Cirrus (17.5 μm [22] and 8.8 μm [59]), and 3D OCT (8.4 μm [22] and 8.1 μm [59]), again confirming that systematic differences existed among SD-OCT devices, and the measurements from different SD-OCT devices should not be considered equivalently.

Significant differences in RNFL thickness measurements among OCT devices were detected, suggesting that OCT measurements from different devices are not directly comparable and meaning that extra care must be taken when comparing measurements among OCT devices. Speculations were made that the following factors accounted for the systematic differences among OCT devices: 1) the intrinsic differences between the TD-OCT and SD-OCT instruments and the hardware discrepancies between SD-OCT devices, 2) the definition of retinal boundaries in segmentation software, 3) the variability in image processing and analysis software, 4) the scan type (circle scan vs. cube scan) and registration area, and 5) the image quality.[56, 57, 64-66] Conversion equations, calculated from regression models [57] or structural equation models (SEMs) [22], were constructed to compensate for the discrepancies among OCT devices and enable direct comparison for measurements from various OCT devices.[22, 57, 66] However, they were either with high standard errors for both intercepts and slope conversion values,[57] or need a larger sample number to be generalized, and thus were not capable of being applied to all the datasets. Kim et al. developed a scan location matching method to align the scan location between TD-OCT's circle scan and SD-OCT's 3D raster cube scans.[73] Their results showed

14

that scan location variability plays an important role in the measurement differences. With the scan location matching method, the measurement variability was successfully reduced and the measurement agreement was improved, however, the systematic differences between TD-OCT and SD-OCT cannot be fully solved without the help of calibration equations.[73]

In summary, systematic differences in RNFL thickness measurements have been recognized as a serious challenge in clinical practice when clinicians intend to compare measurement from different OCT devices or maintain a long-term follow-up for disease management using multiple OCT devices or generations. Calibration equations established to solve the measurement differences were found to be population dependent and not generalizable for every case. To overcome these obstacles, a novel signal normalization method was proposed in this dissertation.

1.5 PROJECT OBJECTIVES

The ultimate goal of this dissertation was to develop a signal normalization method to minimize the signal characteristic discrepancies among multiple OCT devices so that both quantitative and qualitative assessments become directly comparable regardless of the device differences. This would broaden the application of OCT in both clinical and research fields.

The signal normalization method was developed based on the assumption that OCT signals carry similar information from the same biological tissue but show different signal characteristics among OCT devices. The signal differences stemmed from the use of various light sources with different wavelengths, optical designs, and signal processing methods, which cause systematic measurement differences and inconsistent image visualization. We

15

hypothesized that 1) the variation in optical characteristics among OCT devices can be minimized by applying the signal normalization technique; and 2) normalizing the OCT signal will help reduce the inherent systematic differences in quantitative OCT measurement data as well as qualitative OCT image viewing, and enable the direct comparison of both quantitative and qualitative OCT data among multiple OCT devices. To examine these hypotheses, the following objectives were realized:

1.5.1 Objective 1: To Develop and Optimize an Automated Signal Normalization Algorithm

Different OCT devices have different OCT data specifications, such as sampling density in three axes (x, y, and z directions), noise level, and intensity dynamic range. These differences, together with hardware design differences, account for many of the discrepancies in OCT signal characteristics. The signal discrepancies affect the OCT segmentation algorithms, and thus result in systematic differences in OCT measurements. In this objective, we developed a signal normalization method to normalize scaling and sampling density in the axial direction, noise level, intensity dynamic range, and signal strength among various OCT devices in a stepwise fashion. The method was performed in multiple steps and integrated together to form one automated software algorithm.

16

1.5.2 Objective 2: To Test and Validate the Effect of the Signal Normalization Method Quantitatively

In order to test the proposed signal normalization method, quantitative verification of the proposed signal normalization algorithm was performed in two ways: engineering validation and clinical validation. For engineering validation, absolute differences between individual A-scans from different SD-OCT devices sampled at the same location from the same eye were measured and compared. For clinical validation, the circumpapillary RNFL and macular total retinal thickness were measured and analyzed to test the reduction of systematic differences among multiple devices.

1.5.3 Objective 3: To Test and Validate the Effect of the Signal Normalization Method Qualitatively

Even though the OCT signals from different OCT device become similar in shape, for example, fringe patterns, peak widths, and all other aspects, the effect of signal normalization on image appearance still needs to be tested, as subjective qualitative assessment of OCT images is a major part of clinical activity, especially for retinal pathologies. Therefore, a qualitative verification of the proposed signal normalization algorithm was performed subjectively by judging similarity between images from different OCT devices both before and after processing in the following aspects: signal contrast between retinal signal and background noise, signal contrast between adjacent intra-retinal layers, the texture and pattern of OCT images in the retinal layers, and the overall image appearance.

2.0   SUBJECTS AND IMAGE ACQUISITION

In this study, all the experiments conducted in order to evaluate the developed methods were cross-sectional studies. Healthy participants and subjects with various pathologies (glaucoma, AMD, diabetic retinopathy, macular edema, macular hole, and proliferative vitreoretinopathy) were recruited at the University of Pittsburgh Medical Center Eye Center. The University of Pittsburgh Review Board and ethics committee approvals 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.

2.1   SUBJECTS

Healthy subjects, glaucoma subjects, and subjects with retinal pathology were recruited. The inclusion criteria were best-corrected visual acuity of 20/40 or better, refractive error within ± 6.0 D, and no media opacities. Subjects were excluded if they were using medications known to affect the retina, or if they had any previous intraocular surgeries other than uneventful cataract extraction or glaucoma surgery.

2.2   INSTRUMENTS AND IMAGE ACQUISITION

Circumpapillary and macular images were acquired with three SD-OCT devices: Cirrus HD-OCT (Zeiss, Dublin, California, USA), RTVue (Optovue, Fremont, California, USA), and Spectralis (Heidelberg Engineering, Heidelberg, Germany). The details of the scan protocols are described below. For all the images, images with an image quality below the manufacturer's recommended cutoff (signal strength (SS) below 6 for Cirrus data, signal strength index (SSI) below 40 for RTVue data, and image quality below 15 for Spectralis data), 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* images showing a horizontal frame shift larger than one average sized retinal blood vessel diameter or a major distortion of the fovea region. All of the OCT raw data were exported to a standalone computer for signal normalization and further analysis.

2.2.1   Cirrus HD-OCT

Optic Disc Cube 200×200 scan and Macular Cube 200×200 scan were used to obtain the 3D cube data. The scanning protocol collected 200×200 A-scans from a 6×6 mm^2 area centered on the optic disc or the macula with 1024 sampling points within 2.0 mm axial scan depth for each point.

2.2.2 RTVue OCT

200×200 and 513×101 raster cube scan centered on the foveola, and RNFL 3.45 Circle scan pattern were used to acquire RTVue image data at the macula and ONH regions respectively. Isotropic 200×200 raster cube scan patterns collected 200×200 A-scans from a 6×6 mm^2 area centered on the macula and 640 sampling points within 1.96 mm axial scan depth for each point. Anisotropic 513×101 raster cube scan acquires similar information from the same area with the same sampling density in the axial direction, except that 513×101 A-scans are collected. RNFL3.45 Circle scan pattern consisted of 1019 A-scans and 768 samplings along each A-scan for a 2.3 mm axial scan depth following a 3.45 mm diameter circle centered on the ONH.

2.2.3 Spectralis

Macular raster volume scan centered at the fovea covering a 20°×20° region (193 sections, 9 frame averaged), and Circle RNFL scan (100 frame averaged) were used to acquire the image data for the macular and ONH region in this study, where for both scan types 496 sampling points were collected along a 1.9 mm axial scan depth.

3.0 SIGNAL NORMALIZATION ALGORITHM DEVELOPMENTS

Both hardware and software specifications vary significantly among SD-OCT devices; including different light sources, optical design, scan settings, image-processing techniques, and data formats. As summarized in Table 1, Cirrus uses a superluminent diode laser with a center wavelength of 840 nm while Spectralis OCT uses a dual-beam SD-OCT and a confocal laser scanning ophthalmoscope (CSLO) that uses a wavelength of 870 nm and an infrared reference image to simultaneously provide images of ocular microstructures. OCT data can have 1024 pixels along the z-axis in 2.00 mm with an 8-bit data format or 496 pixels in 1.9 mm with a 16-bit data format. All these differences account for the signal characteristic variation among SD-OCT devices, leading to the inherent systematic differences in OCT data measurements. When looking at an individual A-scan at the same location from the same eye scanned by Cirrus and RTVue on the same day, although the signal characteristics varied dramatically in intensity dynamic range and sampling density between devices, the A-scan intensity profiles showed similar profiles in general, such as major peak locations and their relative amplitudes proportions. With these observations, we hypothesized that by normalizing the variable signal characteristics, including sampling density, spatial scaling, and signal to noise ratio, differences among SD-OCT devices would virtually disappear.

We classified factors causing the variability in signal characteristics into four categories: sampling density in the axial direction (Z-direction), the amount of speckle noise, the overall

noise level and intensity dynamic range, and the intensity contrast within tissue signals. We then developed a signal normalization method to resolve each factor in a stepwise fashion. The proposed signal normalization consists of four stages (as shown in Figure 4): z-scaling, sampling density normalization, speckle noise reduction, amplitude normalization, and the optimization of signal to noise ratio using either high dynamic range (HDR) imaging concept based image processing or histogram matching. The details of each stage are described in the following sections. In brief, z-scaling and sampling density normalization resolved the scaling and sampling density variation in the axial direction; speckle noise reduction removed the speckle noise; amplitude normalization rescaled the meaningful retinal signal to the entire intensity dynamic range and normalized the intensity dynamic range; and HDR and histogram matching processing compensated for the image quality discrepancy. All this processing was performed step-by-step. Among our various component specific experiments and validations, the algorithm details of components may exhibit minor differences to suit the purpose (as described in Chapter 4).

Cirrus data format (1024 pixels/samplings in 2.0 mm in the axial direction and an 8-bit data format for one pixel) was used as the normalization reference data format, so that OCT data from other SD-OCT devices were normalized into Cirrus equivalent data format. The reasons for normalization using 8-bit data format as the standard data format can be explained from three aspects. Current display systems are using the 8-bit data format for display purposes; therefore, even though we have higher dynamic range, such as 12-bit or 16-bit data format, the down-sampling of the OCT data is needed for display, which might degrade the image quality. For image processing performance, a majority of the image processing techniques used in computer vision or computer graphics focus on 8-bit data format images, and it has been proved that 8-bit

data format is sufficient to provide satisfactory and reliable processing and segmentation results.[74] Finally, data size is made smaller by converting OCT data to 8-bit data format, which would save memory space and accelerate the processing time of image analysis and data transfer.

Signal Normalization
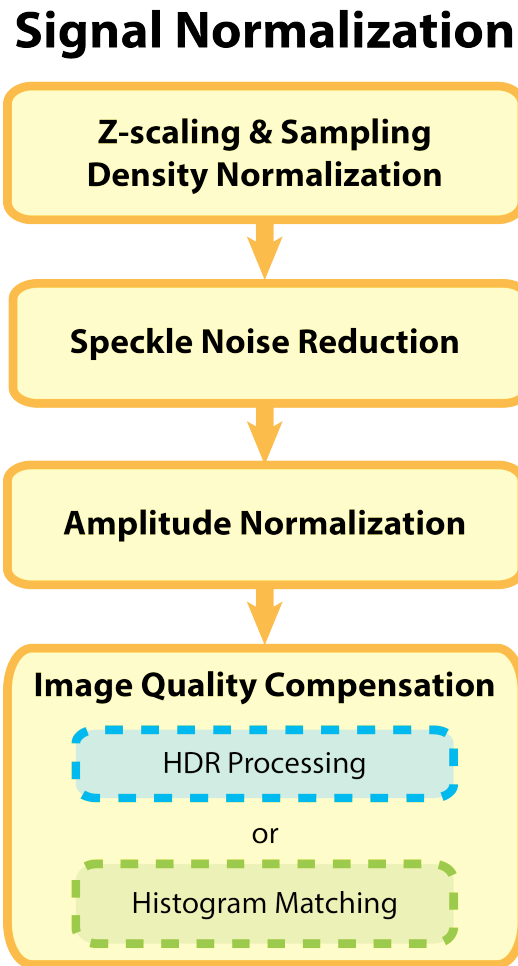
Figure 4. The overall flow of the proposed signal normalization method.

All the image processing techniques used to build the signal normalization method were population independent and did not require training or tuning parameters or converting equations for different OCT devices or cohorts, and could be applied to any OCT devices. We believed the combination of those image-processing methods could successfully minimize OCT signal

discrepancies. Besides, the use of a uniform OCT data format would allow for an established common standard for OCT image data format so one universal visualization and analysis software can handle all OCT data independent of device specific differences.

3.1   Z-SCALING AND SAMPLING DENSITY NORMALIZATION

The simplest way to compare two A-scans is by aligning them to a reference point, such as the ILM or the inner-most border of the retina, and comparing the A-scan intensity profile pixel-by-pixel. However, the discrepancies in z-scaling and sampling density among SD-OCT devices make each sampling point stand for different physical sizes and points to different locations even though they are aligned and with the same index in the axial direction.

As the first step of OCT signal normalization, z-scaling and sampling density normalization were applied. From the device specification, Cirrus data have 1024 sampling points within a 2.00 mm scan depth, RTVue data have 640 sampling points within 1.96 mm for macular scans and 768 sampling points within 2.35 mm for ONH scans, and Spectralis data have 496 sampling points within a 1.90 mm scan depth. As we set Cirrus data format (1024 sampling points within 2.00 mm scan depth; each sampling point stands for about 2 μm) as the reference data format, RTVue and Spectralis data were oversampled along the axial direction using backward mapping to have the same scaling scale and sampling density, as shown in Eq. 3-1:

$$z_{\text{target}} = z \times \frac{ScanDepth_{\text{ref}}}{N_{\text{ref}}} \times \frac{N_{\text{in}}}{ScanDepth_{\text{in}}} \qquad \text{(Eq. 3-1)},$$

where z_{target} indicates the target z index to be sampled for the interpolation, z is the index in the z direction of the pixel in the sampling density normalized A-scan, ranging from 0 to 1023,

24

ScanDepth$_{ref}$ and *ScanDepth*$_{in}$ are the scan depth for the reference data format (2.0 mm), and for the input device (1.96 mm for RTVue macular scans, and 1.90 mm for Spectralis); N_{ref} and N_{in} are the sampling points for the reference data format (1024 pixels), and for the input device (640 pixels for RTVue macular scans, and 496 pixels for Spectralis), respectively. After calculating z_{target}, one-dimensional linear interpolation was applied to generate a smooth interpolated A-scan profile. Sampling points located beyond 1.96 mm scan depth for RTVue data and 1.90 mm scan depth for Spectralis data were padded with the minimum value in the corresponding A-scan.

3.2   SPECKLE NOISE REDUCTION

Speckle is an inherent component of OCT image. When the interference of backscattered waves of the same frequency but with different phases and amplitudes happens, speckles occur, and present as a granular pattern on the OCT cross-sections.[75, 76] Speckle carries information about both the imaged structure (signal-carrying speckle) as well as noise (signal-degrading speckle),[76] and the latter is responsible for the salt-and-pepper appearance especially in areas where less or no signal is expected, such as the vitreous cavity, or the anterior chamber. Speckle noise has been recognized as a factor degrading the quality of acquired images, masking actual target signals, and causing difficulties in qualitative assessment as well as quantitative analysis. We further speculate that speckle noise also results in the variability in OCT signal between OCT images, even within the same device.

Many speckle noise reduction methods have been developed, such as adaptive filtering, anisotropic diffusion, wavelet techniques, etc., and applied as part of pre-processing before

automated segmentation analysis.[77-81] Although these methods can produce good contrast enhancement of the images and maintain the sharpness of the edges, they also generate blurring effects on the fine-textured tissue structure regions, especially within the same retinal layers, which can affect qualitative assessment and potentially influence the automated segmentation performance. Further, adaptive filtering based and anisotropic diffusion based methods require a large number of iterations to complete the process, which is computation-heavy and time-consuming, and thus not practical for real-time clinical diagnosis applications, especially with volumetric scans.[80]

In order to successfully reduce the speckle noise in a time-efficient manner, and keep most of the retinal structural information without blurring the details, a fully automated speckle noise reduction method was developed and tested.

3.2.1 Speckle Noise Reduction Algorithm Development

The speckle noise reduction method was performed as a four-step process (Figure 5):

Step 1: Signal Mapping on Raw Data (Amplitude Normalization)

Each A-scan signal was rescaled based on the pixel intensity histogram (Figure 5, Step 1). In order to maximize the dynamic range within the meaningful retinal signal, signal levels between the 66th percentile and the 99th percentile on the histogram were rescaled onto the 8-bit data grayscale level (lower 66th percentile becomes 0 and top 1st percentile becomes 255). The cutoff of the 66th percentile was determined based on the average thickness of the meaningful signal focusing on the retina (retina with a part of choroid; 660 µm) within the 2.0mm scan length window. This cutoff is commonly used on many devices for the pseudo-color visualization of OCT image data.

Step 2: Tissue Signal Modeling by Smoothing (Generating Tissue Skeleton)

After signal mapping, a two-dimensional (2D) mean filter was applied to smooth the signal on each processed B-scan. Kernel size in the axial direction of the mean filter was a function of the total pixel number in the axial direction. A kernel size of 3×7 (3 in the transverse and 7 in the axial direction) was used for data with 1024 pixels in the axial direction (Figure 5, Step 2).

Step 3: Model Signal Rescaling

After mapping and smoothing, the intensity level of the processed A-scan became lower than the original A-scan profile due to averaging. In order to preserve the original intensity level, the processed A-scan signals were normalized based on the pixel intensity histogram again. The noise level (66th percentile) and saturation level (99th percentile) on the histogram were calculated for the original and the processed A-scan profiles (Figure 5, Step 3). Eq. 3-2 was constructed and used to match the signal intensity level of the processed A-scan to the original A-scan profile.

$$I_R = \frac{I_S - noise_S}{saturation_S - noise_S} \times (saturation_O - noise_O) + noise_S \qquad \text{(Eq. 3-2)},$$

where I_S and I_R refer to the pixel intensity before and after signal rescaling (Step 3), $saturation_O$ and $noise_O$ refer to the saturation and noise level on the original A-scan, and $saturation_S$ and $noise_S$ refer to the saturation and noise level on the smoothed A-scan profile.

Step 4: Selective Removal of High Frequency Noise Components

The rescaled smoothed A-scan signal generated in Step 3 was used as a mask to selectively remove high signal peaks, which were considered as speckle noise. Signal intensity at the same location (with the same z index) from the original A-scan and the mask were compared pixel by pixel (Figure 5, Step 4). If pixels from the original signals showed higher intensity than their counterparts from the mask, their intensity were adjusted to the same intensity as the mask; otherwise, they were kept as the original intensity (Eq. 3-3).

$$I_{SpR} = \begin{cases} I_R & \text{if } I_O > I_R \\ I_O & \text{if } I_O \leq I_R \end{cases} \qquad \text{(Eq. 3-3)}.$$

Eq. 3-3 describes the process of selectively removing the high frequency noise components, where I_{SpR} is the final result, I_R is the pixel intensity from the mask, and I_O is the original pixel intensity.
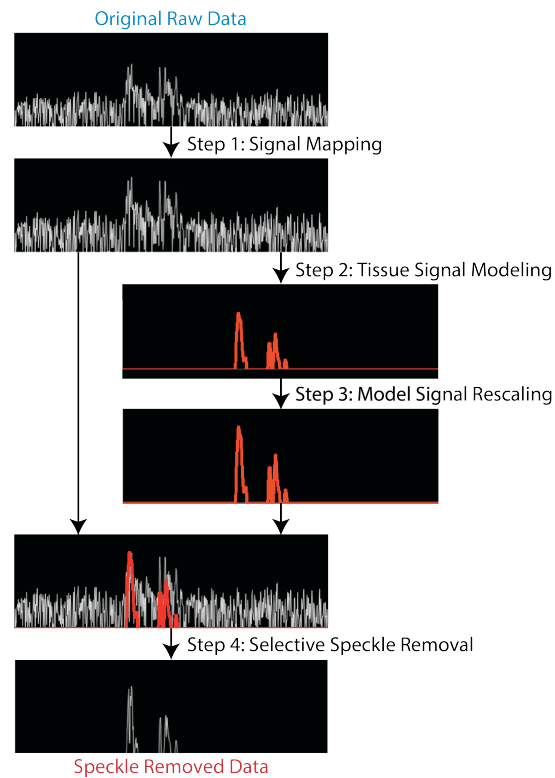
Figure 5. The speckle removal process.

3.2.2 Speckle Noise Reduction Effects Assessment

Subjects and Image Acquisition

To evaluate the noise reduction effects, healthy participants and subjects with various pathologies (glaucoma, AMD, diabetic retinopathy, macular edema, macular hole, and proliferative vitreoretinopathy) were recruited at the University of Pittsburgh Medical Center Eye Center. Circumpapillary and macular images were acquired with Cirrus HD-OCT and RTVue. Details of the scanning protocols were described in Section 2.2.

Noise Reduction Effects Assessment

The effects of the speckle noise reduction method were subjectively evaluated on both cross-sectional images (B-scans) and OCT *en face* images. In OCT en face images, which are created by integrating intensity information along the axial direction, one summed A-scan presents a single pixel in the *en face* image and further constitutes the fundus of the retina.

In addition to subjective evaluation, performance of the proposed speckle noise reduction method was also objectively evaluated using established speckle-reduction performance metrics: signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR).[79, 80, 82-84] SNR measures the signal level of a desired signal to the signal level of the background noise, and CNR measures the difference between the area of image feature and an area of background noise. Image signal quality is better with both higher SNR and CNR values. The definitions for these image quality metrics for a single frame as appearing in the literature are described in Eq. 3-4:[79, 80, 82-84]

$$SNR = 10\log\left(\frac{\max(I^2)}{\sigma_n^2}\right),$$

$$CNR = 10\log\left(\frac{\mu_f - \mu_n}{\sqrt{\sigma_f^2 + \sigma_n^2}}\right)$$

(Eq. 3-4).

In the expression for SNR, *I* represents the logged value from the SD-OCT machine output, and σ_n^2 stands for the variance of the background noise region in the logged value. In CNR formula, μ_f and μ_n indicate the mean value of the selected region of interest and of the same background noise region as in SNR, while σ_f^2 and σ_n^2 stand for the variance of the selected region of interest and of the same background noise region as in SNR.

To measure the SNR and CNR for the entire cube data, we modified Eq. 3-4 into Eq. 3-5:

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

$$CNR = 10\log\left(\sqrt{\frac{\sum_{m=1}^{M}\left(\frac{\mu_f - \mu_n}{\sqrt{\sigma_f^2 + \sigma_n^2}}\right)_m^2}{M}}\right)$$

(Eq. 3-5).

In Eq. 3-5, intermediate SNR and CNR for each frame were calculated and accumulated and 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 set of cube data. For SNR and CNR calculations, signals located in a rectangular region at the bottom of each B-scan, with the same width as the B-scan and height as 3% of the axial pixel number, were considered the noise signal, so that we had the minimal chance to include true retinal signal into the noise signal model when calculating μ_n and σ_n. An additional four regions were automatically located in the

30

RNFL, ganglion cell and inner plexiform layer (GC-IPL), inner nuclear and outer plexiform layer (IN-OPL), and RPE based on the position of ILM, outer RNFL border, and RPE from the segmentation results using our custom segmentation software.[85] The CNR values were averaged over the four regions of interest, while the SNR used the entire image as the region of interest.

For further assessment, we compared the proposed speckle noise reduction method with the conventional mean and median filters by calculating the differences in SNR and CNR before and after processing using our method, mean, and median filters. The noise reduction ability for OCT images from different OCT devices was also tested.

Statistical Analysis

Paired t-tests were used to analyze the image quality metrics (SNR and CNR) improvements between the original images and speckle noise reduced images. Differences in SNR and CNR before and after the noise reduction processing using different methods were calculated. The differences were used as the parameters for paired t-tests to further compare the noise reduction ability in terms of image quality among different noise reduction methods.

3.2.3 Results

Two hundred sixty-nine scans on 155 eyes from 95 subjects were recruited. All images showed notable improvement in image/signal quality, regardless of the scanning location (macula or ONH), pathology, or signal strength. Speckle reduced *en face* images showed noticeably clearer borders of retinal vessels and ONH contours compared to the raw image. Speckle reduced cross-sectional images exhibited minimal blurring while speckle noise was effectively suppressed and

edges between different layers were preserved. A-scan profiles displayed clear retinal structure signal. Average processing time for each scan (200×200 cube data) was 4.3 seconds (MacBook Pro, 2.6GHz Core i7, 8GM RAM).

Quantitatively, both SNR and CNR showed statistically significant improvement after speckle noise reduction regardless of the methods (p<0.0001, paired t-test) (Table 2). Compared to the conventional mean and median filters, the present method outperformed both conventional methods (p<0.0001), except for the CNR on Cirrus images.

Table 2. Image quality metrics using different noise reduction methods. SNR: signal-to-noise ratio, CNR: contrast-to-noise ratio, and Diff: SNR or CNR difference from original to each noise reduction method.

		SNR	**Diff**	**CNR**	**Diff**
Cirrus (N=124)	Original	26.2	---	4.3	---
	Mean Filter	29.1	2.9	5.6	1.4
	Median Filter	28.8	2.6	5.4	1.2
	Our Method	47.0	20.8	5.0	0.7
RTVue (N=145)	Original	20.8	---	2.7	---
	Mean Filter	27.1	6.3	5.3	2.6
	Median Filter	26.2	5.4	4.9	2.2
	Our Method	43.8	23.0	5.8	3.1

For qualitative assessment, five sample cases are shown in Figure 6 to illustrate the effect of the speckle reduction.

	OCT *en face* image	Cross-sectional OCT image (False color version)	A-scan

Clinical Display

Speckle Reduced Data

Case 1

	OCT *en face* image	Cross-sectional OCT image (False Color Version)	A-scan

Clinical Display

Speckle Reduced Data

Case 2

33

| OCT *en face* image | Cross-sectional OCT image (False Color Version) | A-scan |

Case 3

| OCT *en face* image | Cross-sectional OCT image (False Color Version) | A-scan |
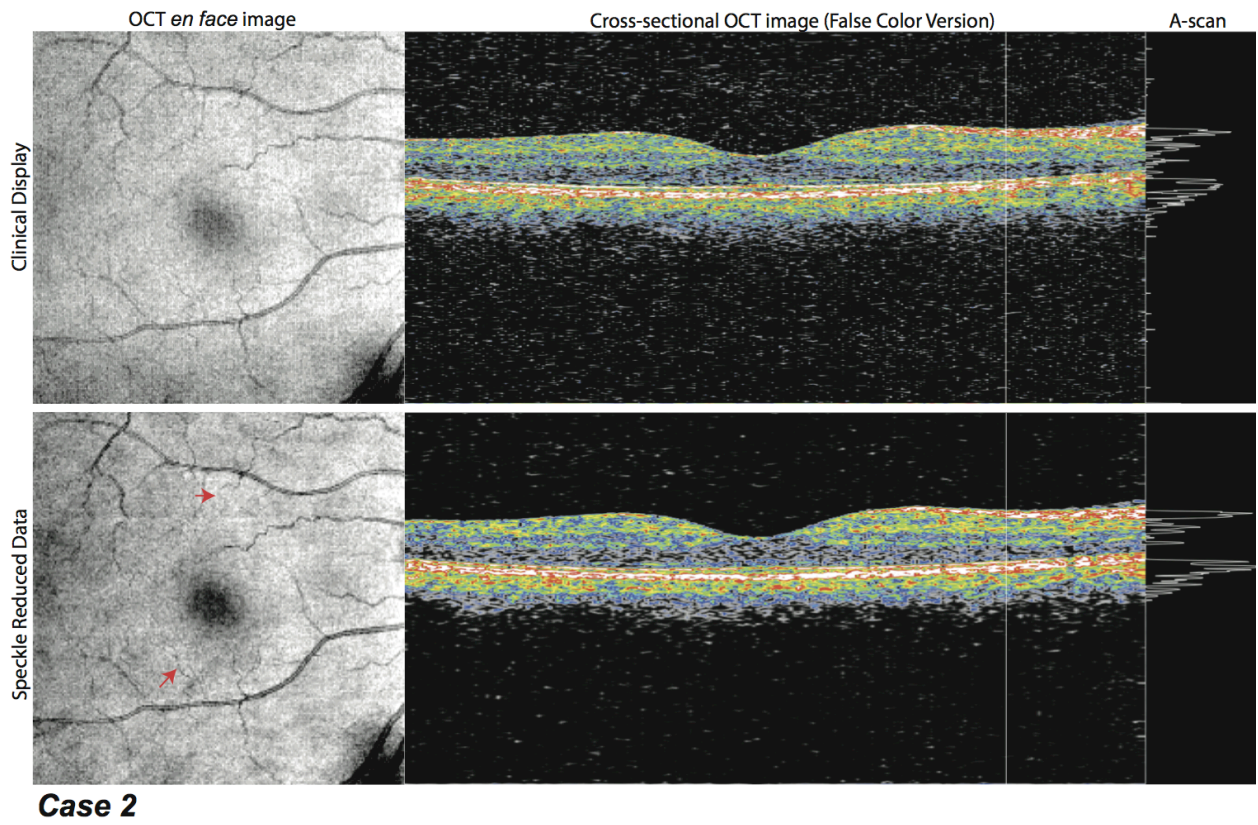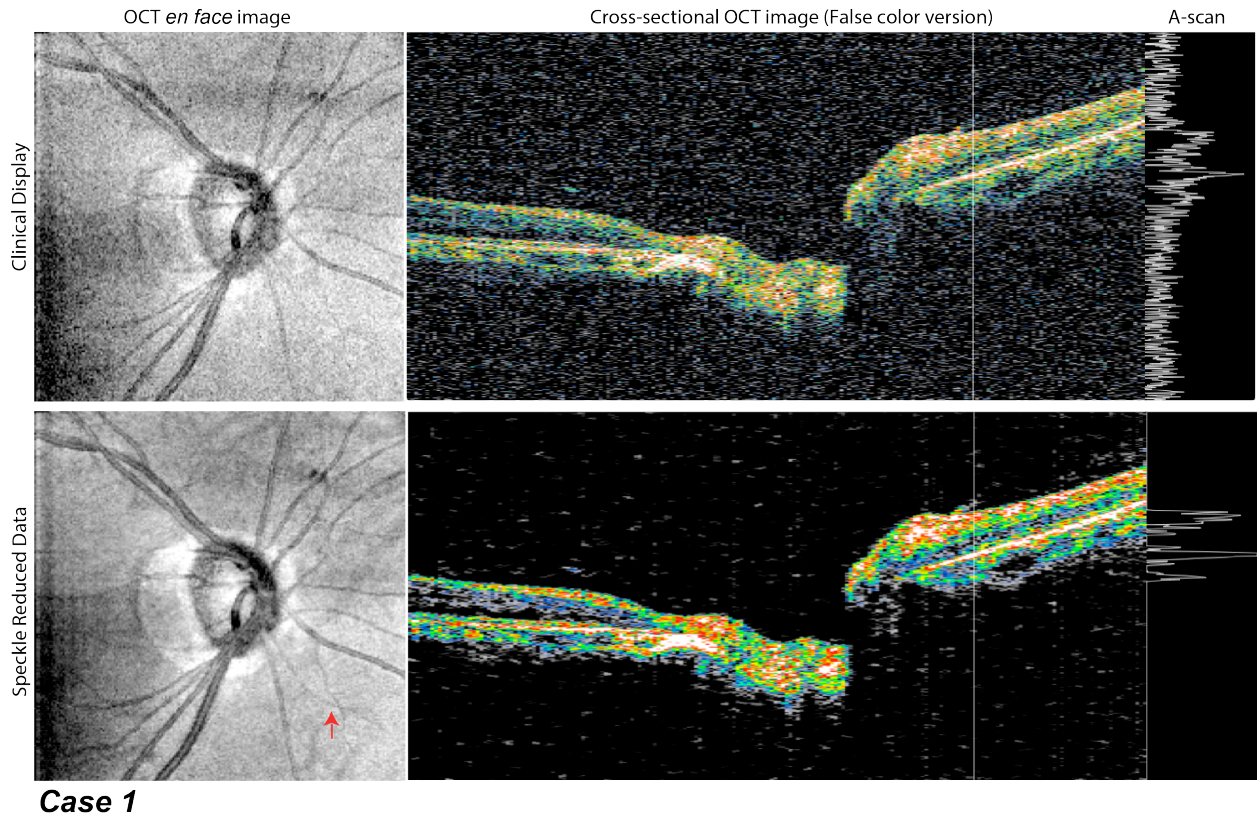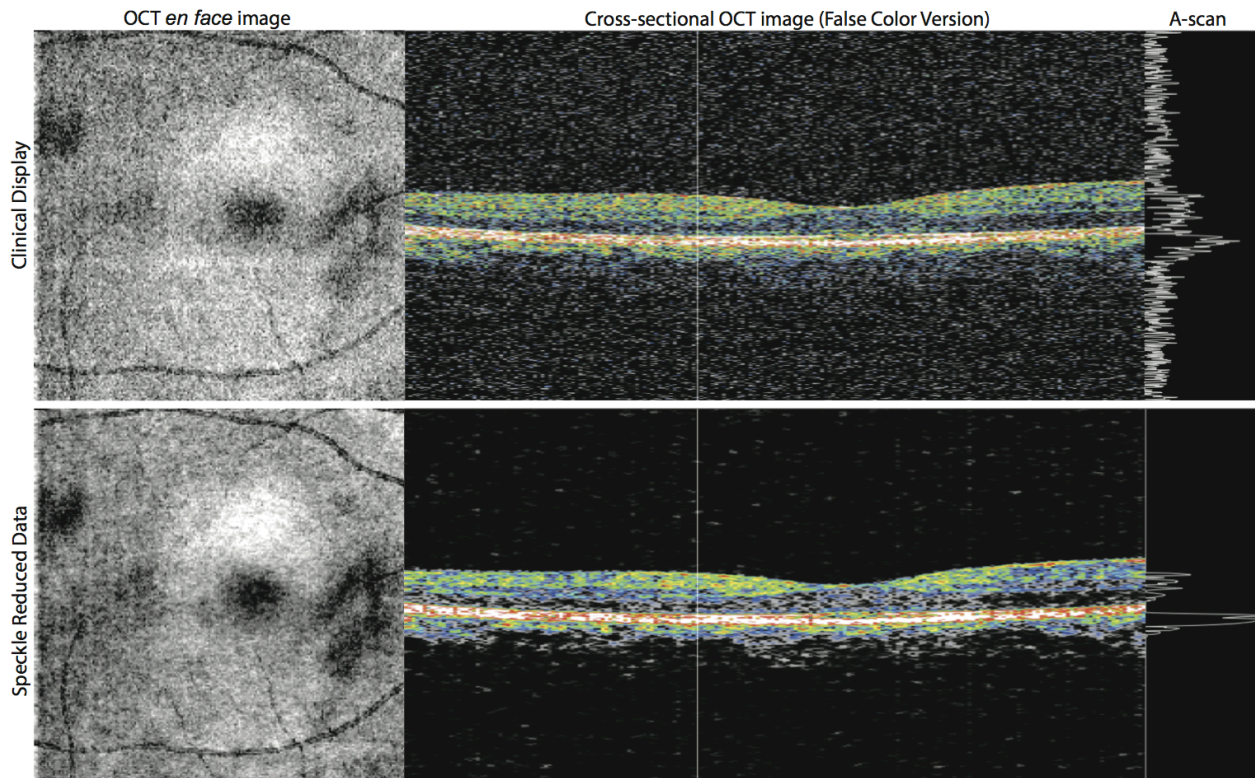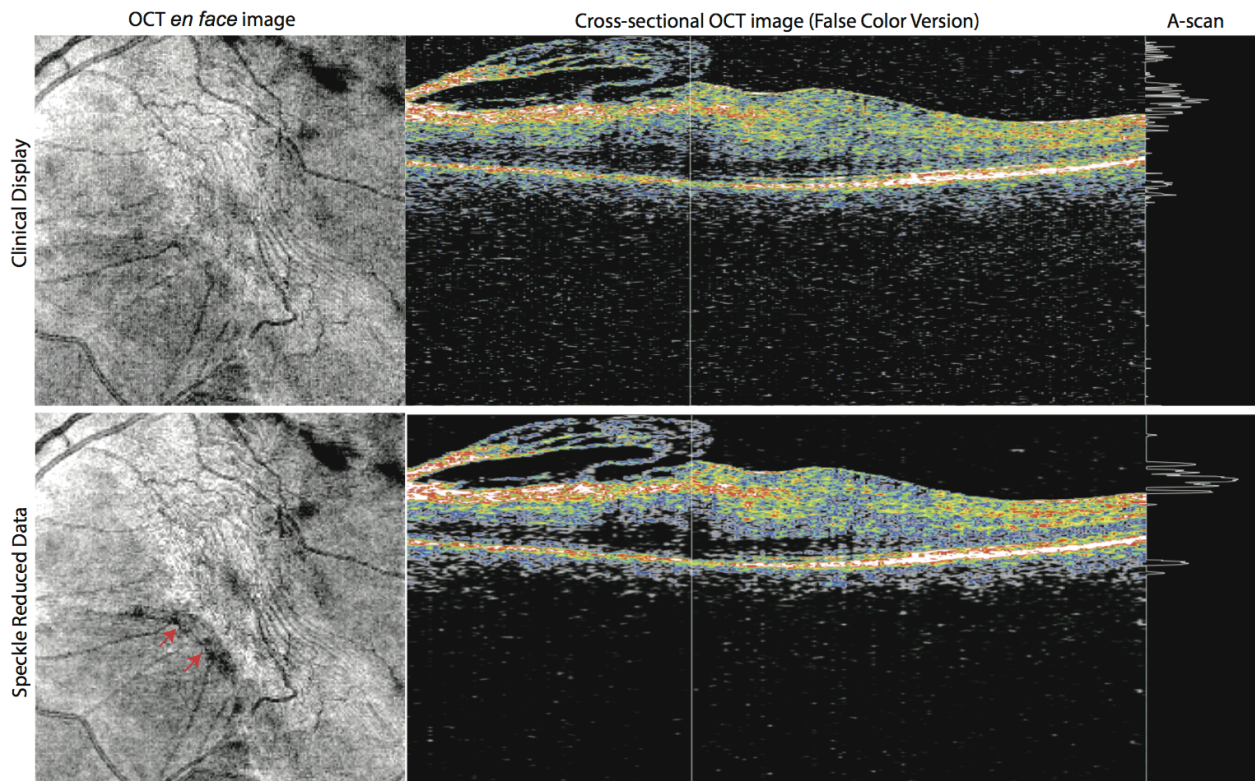
Case 4
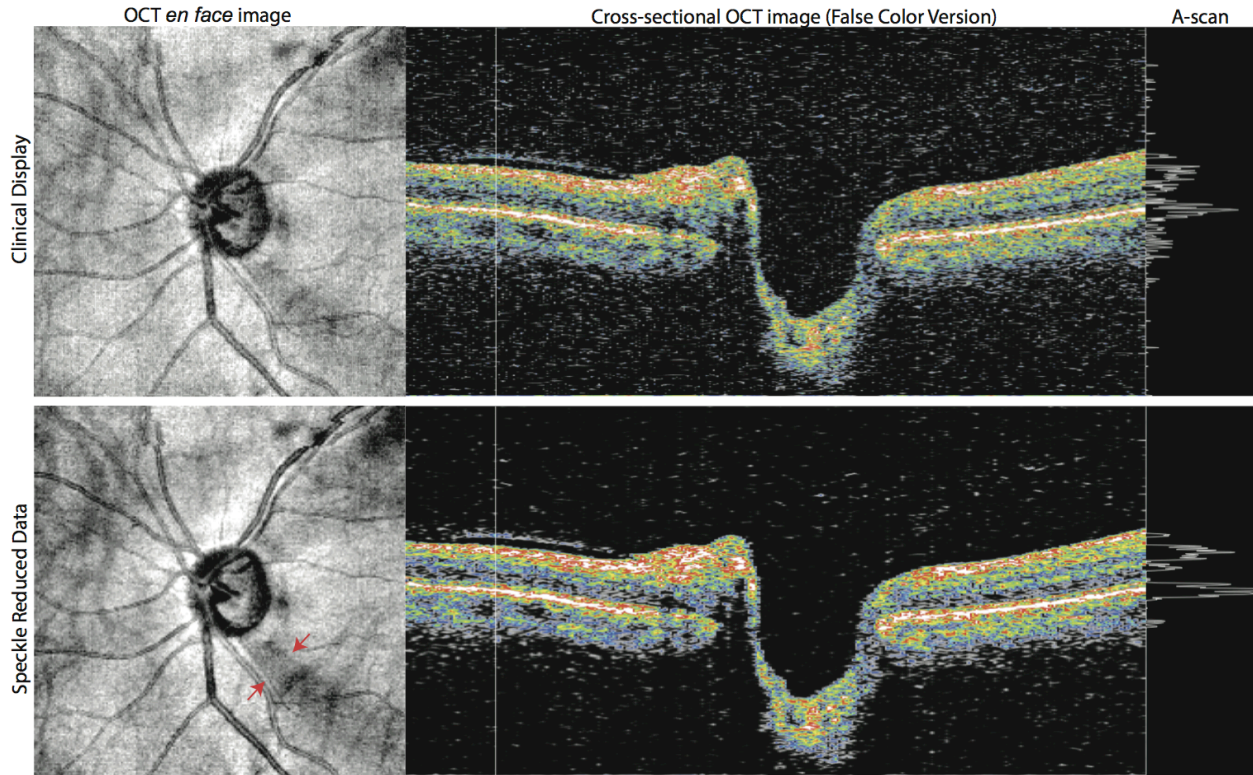
Figure 6. Example results of the presented speckle reduction method. The original en face image (top left) is shown with the cross-sectional B-scan (top middle) and the A-scan profile corresponding to the vertical white line on the original cross-sectional B-scan (top right). The processed en face image (bottom left) is shown with the cross-sectional B-scan (bottom middle), and the A-scan profile corresponding to the vertical white line on the processed cross-sectional B-scan (bottom right).

Case 1

The original OCT data, presented in the top half of Figure 6, Case 1, contained a significant amount of speckle noise, which can be seen as the granular pattern in the *en face* image (left), the salt-and-pepper noise in the cross-sectional B-scan (middle), and the fluctuated shape of A-scan profile (right). After noise reduction processing, speckle noise was reduced both in the cross-sectional image (bottom middle) and the A-scan intensity profile (bottom right). Retinal layer boundaries became clearly visible and the contrast was enhanced. In the *en face*

image, the grainy appearance was cleaned up and the borders of blood vessels and ONH became sharper (shown in the bottom left figure). Even small vessels became noticeably visible after processing (as the red arrow indicates).

Case 2

Although there was little speckle noise visible in the original OCT data of Case 2 (Figure 6, Case 2), both the cross-sectional image and the A-scan profile became cleaner after processing. The *en face* image also showed noticeably less noise, and small blood vessels became noticeably more visible after processing.

Case 3

In Figure 6, Case 3, OCT data not only contained notable speckle noise but also had poor signal quality, likely due to poor focus. Although speckle noise was reduced and the contrast was enhanced on the cross-sectional image and the A-scan profile, the speckle reduction effect was limited in the *en face* image due to poor focus.

Case 4

In Case 4 (Figure 6, Case 4), vitreo-retinal proliferative change is seen in both the cross-sectional and *en face* images. Speckle was reduced in all 3 visualizations (*en face*, cross-section, and A-scan). Visibility of proliferative tissue insertion in the retina was increased with processing (red arrows). Contrast of retinal folds was also improved.

<u>Case 5</u>

A glaucomatous eye with retinal nerve fiber wedge defects is shown in Figure 6, Case 5. After processing, the quality of all visualizations (*en face*, cross-section, and A-scan) was improved and the wedge defect (red arrows) was also effectively highlighted. Image contrast was enhanced and the visualization of the retinal nerve fiber wedge defect was improved.

3.2.4 Discussion

The presented speckle method improved image quality in both quantitative and qualitative assessments with little blurring. The innovation of the presented method is the selective removal of the high frequency noise components. Unlike existing noise reduction methods, which remove all the high frequency components without considering the intensity level to generate smooth areas when the local intensity variance is small, the presented noise reduction method only removes the high frequency components with high intensity and keeps the high frequency components with low intensity as original. By doing this, the output signals maintain details in the low and mid intensity level, which are recognized as tissue structures, and thus prevent the processed images from being blurry, solving a common side effect of most of the speckle noise reduction methods.

The selective smoothing approach can be applied to any other noise reduction method where signal modeling takes place. One of the common limitations of noise reduction is image blurring. Blurring happens because the original signal is modified from both ends: high and low signal level. For a majority of the cases, unwanted speckle noise belongs to a high signal level

not low (salt and pepper appearance), especially for subjective image quality assessment. We believe that the selective approach is a simple but powerful concept in practical signal enhancement.

Besides the preservation of tissue structure information, another advantage of the presented method is time-efficiency. Our noise reduction method removes the noise by first generating a smoothed mask and then comparing the original signal with the mask and selectively smoothing. In this way, the computational complexity only relates to the number of A-scans and frames, substantially reducing the computation complexity of the existing methods (such as a complex statistical model or the need for a large number of iterations). This is especially important as the number of sampling points keep increasing with the advancement of OCT sampling speed.

Although the presented noise reduction method can remove speckles and enhance the image contrast and quality, inferior quality images due to poor focus, polarization, or factors not related to speckle noise showed limited improvement with the presented method. Poorly focused images present with a deficiency in signal content. Therefore, even after removing high frequency noise components, these images did not show the full details of the actual tissue, leading to less optimal visualization.

For objective image quality evaluation, the overall SNR and CNR showed a significant improvement in the image quality in terms of signal strength and image contrast. However, we found some systematic difference in the magnitude of improvement in CNR among SD-OCT devices (Table 2). For Cirrus data, the CNR increased from 4.3 dB to 5.0 dB, while for RTVue data, the CNR increased from 2.7 dB to 5.8 dB, regardless of the scan location and scan type. This phenomenon can be explained by the definition of CNR. When calculating CNR, the mean

of the region of interest is subtracted from the mean of the noise region and then divided by the averaged standard deviation (SD) of the region of interest and noise region. The original RTVue data have a higher noise level and a larger SD as compared to Cirrus data as indicated in the red boxes in Figure 7, and therefore provides a large improvement of the CNR after speckle noise reduction.
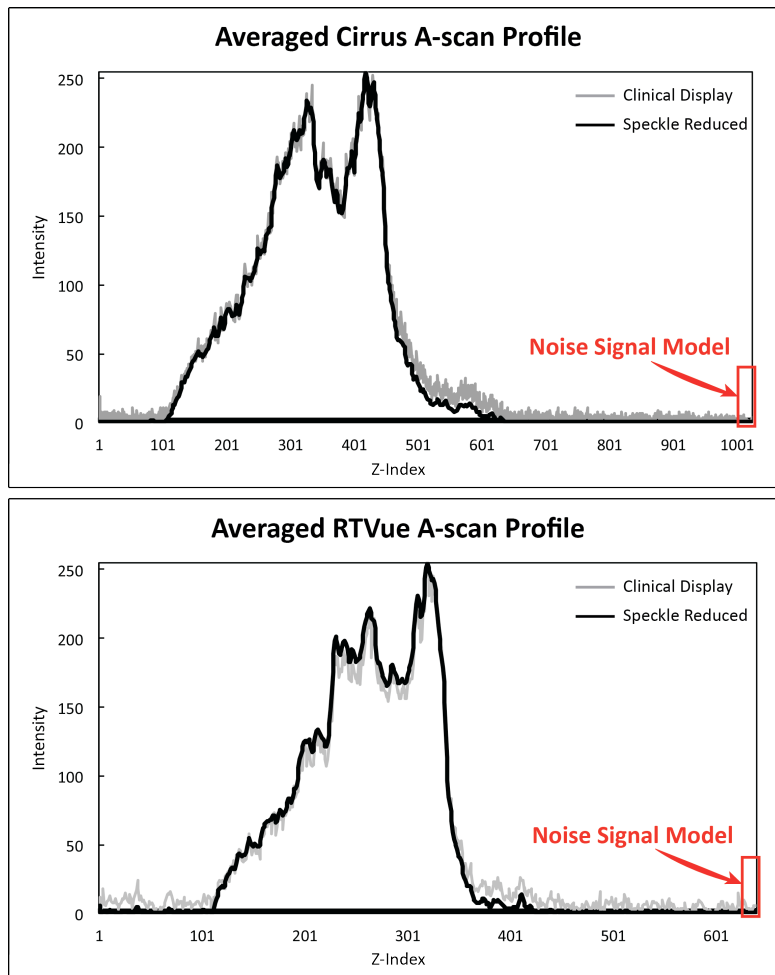
Figure 7. Averaged A-scan before and after speckle reduction with two SD-OCT devices. Averaged A-scan profiles in one single frame before and after speckle noise reduction from the same eye scanned on the same day using Cirrus and RTVue machines. The clinical display and the speckle reduced A-scan profiles are superimposed together to better observe the effect. Red boxes indicate the regions used as noise signal models when calculating SNR and CNR for Cirrus and RTVue data.

CNR results also showed contradicting trends between Cirrus and RTVue images. The present method outperformed the conventional mean and median filters on RTVue but not on Cirrus. In theory, CNR gives an advantage to the conventional methods because mean intensity is higher and variance is lower with simple smoothing than the present method, as the present method preserves high frequency components at low intensity level. However, RTVue images have higher noise floor level than Cirrus images (Figure 7), and the conventional methods do not suppress the noise as effectively as the present method. Therefore, the background noise effect overwhelmed the other factors, which depend on the region of interest.

In conclusion, the presented noise reduction method successfully reduced speckle noise, enhanced image contrast and quality, and preserved the details of tissue structure in a time-efficient manner. It was effective across all tested OCT platforms, which may form the foundation of a clinically useful post-processing tool.

3.3 HIGH DYNAMIC RANGE IMAGING CONCEPT BASED SIGNAL ENHANCEMENT METHOD

It is well known that image quality variability affects our ability to interpret and analyze OCT images.[70, 86] Quantitative RNFL thickness measurements showed a significant positive correlation with the image quality.[70, 72, 86] Qualitative evaluation of OCT images is also markedly influenced by the signal quality.[87-89] Several attempts were designed to enhance OCT images with relatively low image quality,[69, 90-92] however, to our knowledge, there is none that proved to be effective in addressing OCT measurement variability within the same target due to variable signal quality.

HDR imaging technology, which has a long history in photography, expands the image contrast dynamic range by combining multiple images with a wide range of exposure settings.[93-98] Figure 8 shows an example of the effect of HDR. Though, with OCT scanning, acquiring multiple scans with different exposure settings is not feasible, as the exposure level cannot be controlled mechanically or optically. We hypothesized that the HDR concept can be applied to enhance OCT images and achieve greater dynamic range in both weak and strong signal areas without the need of multiple scans, and that the HDR processing technique can be used to compensate image quality differences in quantitative and qualitative OCT image assessment.
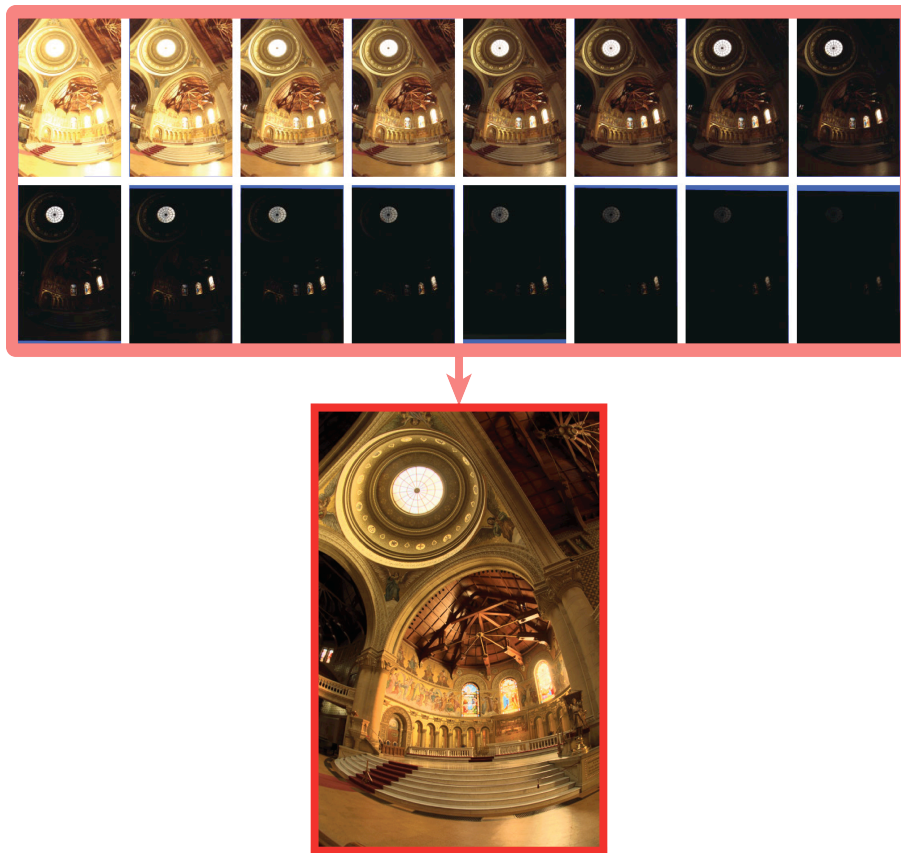
Figure 8. An example result of HDR processing in photography. Using images with multiple exposure setting to generate a HDR image. (Debevec PE and Malik J [93])

The purpose of this section was to develop and test a novel signal enhancement method for OCT images based on the HDR processing concept without the need of multiple scans in different exposures. For its validation, a set of OCT images obtained under varying corneal dryness conditions causing fluctuating image quality were processed to see the effect on RNFL thickness measurements between good and poor signal strength images scanned on the same eye.

3.3.1 High Dynamic Range Processing Algorithm Development

The HDR processing includes two major stages:

Step1: Three Virtual OCT Signal Channels Processing

For each B-scan image, four histogram parameters, minimum, maximum, noise level, and saturation level were calculated based on a previous study,[87] where minimum and maximum were the lowest and highest pixel values of the entire B-scan image, while noise level and saturation level were defined as the 66th and 99th percentile of the pixel value of the entire B-scan. For each frame, the original OCT signal dataset was divided into three datasets, creating three virtual channels: low, medium, and high signal channels. The low signal channel, I_{Low}, consisted of pixel values between minimum and low offset values, the high signal channel, I_{High}, consisted of pixel values between high offset and saturation level values, and the medium signal, I_{Mid}, consisted of pixel values between low offset and high offset values, where low and high offsets were defined as Eq. 3-6:

$$Low\ Offset = noise + 0.23 \times (saturation - noise),$$
$$High\ Offset = saturation - 0.067 \times (saturation - noise)$$

(Eq. 3-6).

42

Each dataset is then processed to maximize the signal dynamic range by linearly rescaling pixel values between lowest and highest values within each dataset to the full 8-bit grayscale range (0 to 255) in each B-scan. Intensity values outside of the defined cutoff values (lower or higher) are forced to be either 0 or 255.

Step 2: High Dynamic Range Signal Composition

Signals from all three channels are combined to generate the final HDR dataset by calculating weighted mean values of the three channels, as shown in Eq. 3-7, where I_{Low}, I_{Mid}, I_{High}, and I_{HDR} stand for low, medium, and high signal channels, and the output image after HDR processing, respectively; $I(x,z)$ indicates the pixel value at position (x,z) in the processed B-scan, i.e. x^{th} A-scan and z^{th} pixel in the axial direction; and c_L, c_M, and c_H are the weighted coefficients. The coefficients used for calculating the weighted mean are adjusted so that the image quality can be enhanced for OCT images with poor signal strength, while preventing the images with good signal strength from becoming saturated. In general, the coefficients were 3.0 for c_L, 2.0 for c_M, and 1.0 for c_H.

$$I_{HDR}(x,z) = \frac{1}{c_L + c_M + c_H}\left(c_L \times I_{Low}(x,z) + c_M \times I_{Mid}(x,z) + c_H \times I_{High}(x,z)\right) \qquad \text{(Eq. 3-7)}.$$

3.3.2 High Dynamic Range Processing Performance Assessment

Two experiments were designed to test the performance of the proposed HDR processing on OCT images. In the first experiment, we tested the signal quality compensation ability of the HDR processing using our previous TD-OCT data, which had a wide range of SSs.[70] In the

second experiment, we moved forward onto SD-OCT data to assess the image appearance after HDR processing, and to show that the HDR processing also works on SD-OCT data.

Experiment 1: Effects of Signal Quality Compensation on Quantitative Analysis

This experiment was designed to test the effects of the proposed HDR processing method on signal quality compensation for quantitative analysis. The variability in RNFL thickness measurements on images scanned with various SS of the same eye was measured.

<u>Subjects and Image Acquisition</u>

Ocular images obtained in a previous study were used in this experiment.[70] Seventeen eyes of 17 healthy volunteers were scanned with the TD-OCT Fast RNFL scanning protocol (Stratus OCT), which generated 3 consecutive, circumpapillary RNFL images at a scanning radius of 3.4 mm centered on the ONH. The upper eyelid of each subject was taped to the forehead on the selected, anesthetized eye to prevent blinking so that images with a wide variety of signal quality could be acquired from the same eye. In general, corneal dryness correlates well with signal quality (the drier the cornea, the worse the signal quality). OCT images were acquired every 20 seconds for a total of 8 series of images on each eye. After the drying scans, the tape was removed and the subject was allowed to blink normally. Then, 3 more scans were acquired at 1, 2, and 4 minutes after removing the tape. In this way, we had the reference scans (the scans with the highest SS) and the deteriorated scans all acquired from the same eye and same session, removing most potential confounders. All of the raw image data were exported to a standalone computer for further HDR processing.

Both the original OCT data and the HDR processed OCT data were then processed with our custom segmentation algorithm to measure RNFL thickness. The details of the segmentation algorithm are described in another study.[85] For the original OCT data, RNFL thickness was also measured with the original built-in segmentation algorithm of the machine. Segmentation results were subjectively evaluated for the accuracy of the automated RNFL border detection. Scans were excluded if the images demonstrated one or both of the following: (1) apparently inaccurate border detection for more than a consecutive 15% or additive 20% of the total image or (2) the borders of the RNFL were collapsed, meaning that the RNFL thickness was recorded as a string of zeros for at least 10 consecutive points. The mean of at least 2 qualified scans (from the 3 scans acquired in each series) for each time point was used for the analysis.

Finally, a pair of scans, which were the scans with the highest and lowest signal quality without RNFL segmentation failure were selected for each eye in order to compare the HDR processing effect.

Experiment 2: Effects of Image Quality Enhancement

This experiment was designed to test the effects of the proposed HDR processing method on OCT image enhancement. OCT images with poor signal quality were processed with our HDR processing method, and the visibility of the intra-retinal layers before and after HDR processing was subjectively evaluated. In addition, the objective assessment of image quality was further applied to a separate dataset of SD-OCT images that showed good signal quality in order to test the effects of the proposed method on normal, clinically acceptable images.

Image Enhancement Ability Assessment

Fifteen eyes of 15 subjects (8 healthy and 7 glaucoma) were enrolled. Poor quality images obtained with two SD-OCT devices were assembled for this study (Cirrus HD-OCT and RTVue). The visibility of intra-retinal layers was subjectively evaluated before and after HDR processing by presenting the images in a random order. The observer judged if there was a notable difference in visibility between each pair. In addition, the same custom retinal segmentation algorithm as the one used in *Experiment 1* in Section 3.3.2 was also performed on both original and HDR processed OCT images to test the possible improvement on segmentation performance after HDR processing.

Objective Image Quality Assessment

Two hundred seventy SD-OCT images were acquired from one hundred thirty-six eyes from 95 subjects (32 healthy, 22 glaucoma suspect, and 41 glaucoma subjects). High quality images, which were eligible for clinical diagnoses and image analyses, were obtained with two SD-OCT devices for this experiment (Cirrus HD-OCT and RTVue).

SNR and CNR were used to objectively evaluate the performance of the proposed HDR processing of 270 SD-OCT images. The definition of SNR and CNR were identical to the one used in "Noise Reduction Effects Assessment" (Section 3.2.2). As HDR processed images are clipped at the saturation level (the 99th percentile), SNR and CNR calculation was done on the original images after the same clipping was applied. This prevents erroneous measurements due to a few exceptionally high intensity outlier pixels.

<u>Statistical Analysis</u>

The relationship between RNFL thickness measurements and the SS of OCT images was measured using linear regression models for each of the following RNFL thickness segmentation algorithms: the original device's built-in algorithm (Device) and our custom algorithm with or without HDR processing (Original or HDR, respectively). Paired t-tests were used to analyze the image quality metrics (SNR and CNR) improvement between the original and HDR processed images.

3.3.3   Results

Experiment 1: Effects of Signal Quality Compensation on Quantitative Analysis

In total, 951 images were collected, and 109 (11.5%) of them (mean SS 2.7 ± 1.1) were excluded from the study due to the segmentation failure. The overall mean SS of the original TD-OCT images was 6.6 ± 2.4, ranging from 0 to 10 (the full range of the SS for the TD-OCT device). Mean SS of good and poor quality scans were 9.0 ± 1.1 and 4.4 ± 0.9, respectively. Prior to signal quality enhancement, the RNFL thickness showed significant differences between good and poor quality scans on the same eye (mean RNFL difference 11.9 ± 1.6 μm, p<0.0001, paired t-test). This difference became substantially smaller and non-significant after HDR processing, where the mean difference was 1.7 ± 1.7 μm (p=0.33). This result fits well within the expected test-retest measurement variability, which is 10 μm for TD-OCT images with good signal quality.[99]

The RNFL thickness showed two separate linear relationships with a SS above or below 4 for all algorithms (Device, Original, and HDR, Figure 9). In the range of SS > 4, the HDR algorithm showed a smaller slope (-0.01 μm) in the relationship between RNFL thickness and SS

47

as compared to other algorithms (2.6 μm and 7.4 μm, for Device and Original, respectively). In the range of SS ≤ 4, all algorithms showed similar and relatively steep slopes, ranging from 7.6 to 10.5 μm, in the relationship between RNFL thickness and SS (Table 3).
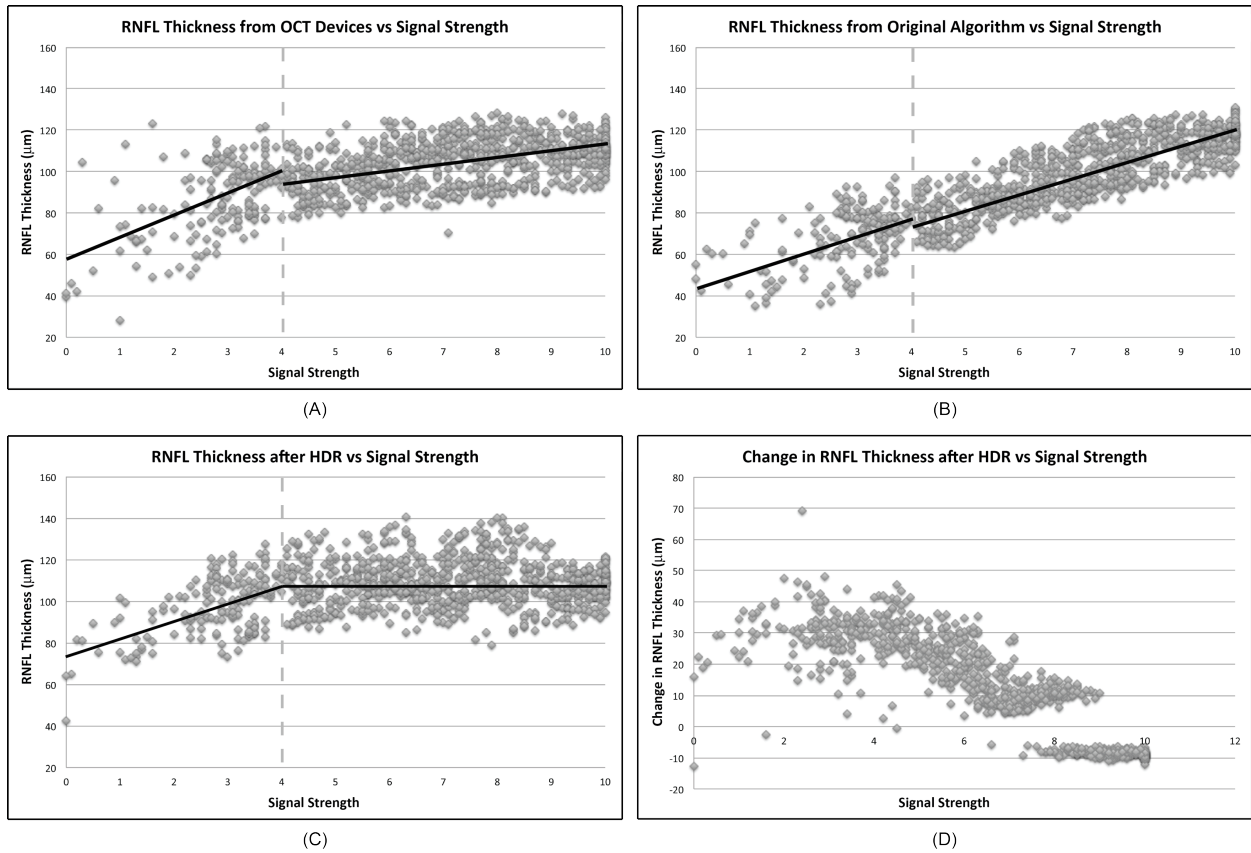
Figure 9. Scatterplots of RNFL thickness measurements using various methods versus SS. (A) The built-in algorithm in TD-OCT devices, (B) our custom segmentation algorithm, and (C) our custom segmentation algorithm after HDR processing. The RNFL thickness showed two separate linear relationships with SS less than or equal to 4 and SS greater than 4 for all algorithms. (D) The changes in RNFL thickness after HDR imaging as a function of baseline SS.

Table 3. Summary of signal quality compensation on quantitative analysis. RNFL thickness (RNFLT) measurements, slope, and intercepts of the regression curves of RNFL thickness versus SS using different processing methods. 95% confidence interval (CI) of the RNFL thickness measurement is shown in the parentheses.

	SS ≤ 4			SS > 4		
	Device	**Original**	**HDR**	**Device**	**Original**	**HDR**
RNFLT (μm)	86.8 (84.0, 89.7)	68.0 (65.9, 70.2)	97.8 (95.6, 100.0)	104.8 (104.0, 105.5)	100.4 (99.3, 101.5)	109.5 (108.7, 110.3)
Slope (μm/SS)	10.5	7.6	9.0	2.6	7.4	-0.01
Intercept (μm)	57.8	46.9	72.9	85.4	46.3	109.6

Figure 9D shows the changes in RNFL thicknesses as a function of the corresponding baseline SS. For lower range SS (SS < 7), there was a significant increase in RNFL thickness after HDR processing (mean thickness change 24.5 ± 10.0 μm), and the amount of change decreased as SS increased, while less changes in RNFL thicknesses were found in higher range SS (SS ≥ 7), with mean thickness change -0.25 ± 9.6 μm.

Experiment 2: Effects of Image Quality Enhancement

In all SD-OCT images with poor signal quality, notable improvements in terms of retinal layer visibility were observed (Figure 10). The contrast between adjacent retinal layers or between layers with high and low reflectivity became more apparent. Areas with poor signal quality that led to segmentation failure were accurately segmented after HDR processing (Figure 10). Seven out of 9 SD-OCT images with segmentation algorithm failure (77.8%) showed successful segmentation after HDR processing.
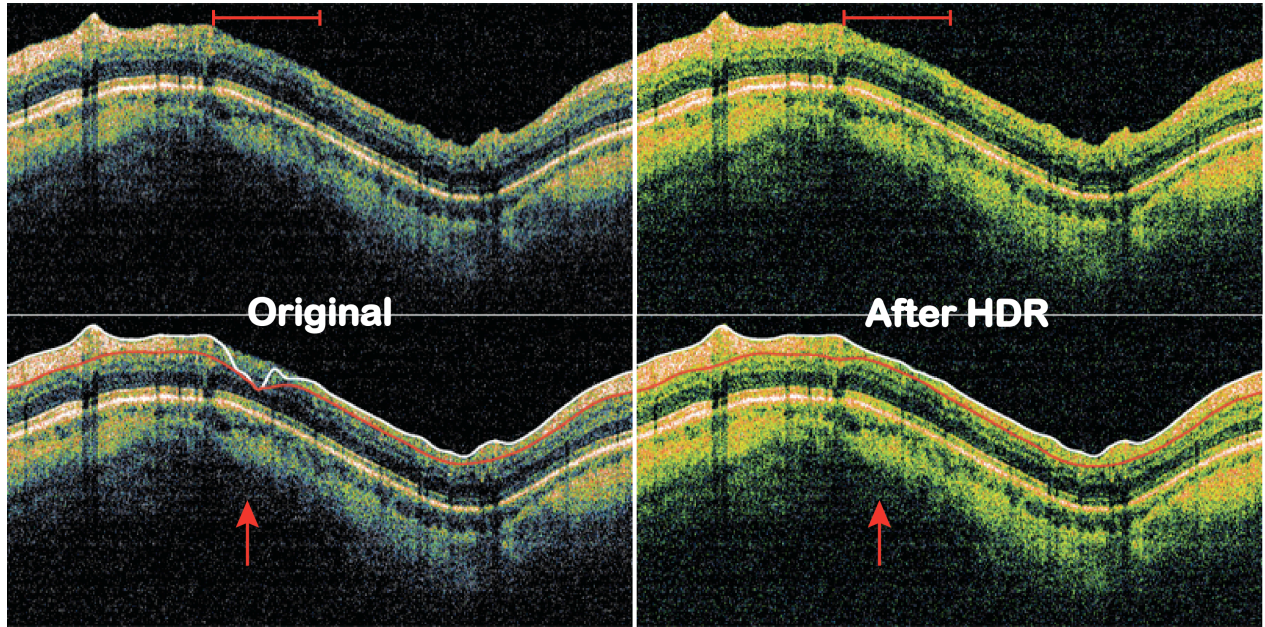
Figure 10. SD-OCT cross-sectional images before and after HDR processing. Top row: Visibility of the retinal layers became clearer across the image, especially the area within the red bar on top. Signal quality also became more homogeneous with HDR processing. Bottom row: RNFL segmentation failed on original image but succeeded after HDR processing (red arrow).

For objective assessment using the image quality metrics, the overall SNR of the processed images was statistically significantly lower than the original SNR, while the overall CNR of the processed images showed statistically significant improvement compared to the original (SNR: 23.3 vs 20.0 dB, CNR: 2.8 vs 3.0 dB; $p<0.0001$, paired t-test) (Table 4). A similar trend was found when analyzing the change in image quality metrics for each imaging device separately. SNR showed a statistically significant decrease for both Cirrus and RTVue data (26.2 vs 20.9 dB for Cirrus, 20.8 vs 19.1 dB for RTVue, $p<0.0001$ for both devices, paired t-test) while CNR showed a statistically significant improvement (2.8 vs 3.1 dB for Cirrus, 2.7 vs 2.9 dB for RTVue, $p<0.0001$ for both device, paired t-test).

Table 4. Image quality assessment results for original and HDR processed images. Diff: SNR or CNR difference from the original to HDR.

	SNR (dB)			CNR (dB)		
	Original	**HDR**	**Diff**	**Original**	**HDR**	**Diff**
Cirrus (N=124)	26.20 (26.12, 26.28)	20.94 (20.85, 21.03)	-5.27 (-5.30, -5.24)	2.80 (2.59, 3.01)	3.07 (2.80, 3.34)	0.27 (0.20, 0.33)
RTVue (N=145)	20.81 (20.75, 20.88)	19.12 (19.10, 19.17)	-1.69 (-1.72, -1.66)	2.73 (2.64, 2.82)	2.90 (2.81, 2.99)	0.17 (0.16, 0.18)
Overall (N=269)	23.30 (22.97, 23.63)	19.96 (19.84, 20.10)	-3.34 (-3.56, -3.12)	2.76 (2.66, 2.87)	2.98 (2.85, 3.11)	0.21 (0.18, 0.24)

3.3.4 Discussion

The proposed HDR processing method successfully compensated for the signal quality variation, reduced the consequent RNFL thickness measurement variability, and minimized the measurement variability across a wide range of SS to the level of expected measurement variability within the good SS range. Most retinal layer segmentation algorithms detect the retinal layer boundaries based on the contrast between the adjacent retinal layers. In OCT images with poor signal quality, the contrast between the adjacent retinal layers is degraded because of low signal quality and loss of tissue information. This leads to variable border detection accuracy. The HDR processing enhanced the image quality in areas with poor signal quality, resulting in the reduction of the RNFL thickness measurement variability.

The HDR method worked well for images with moderately low signal quality (SS > 4). For images with extremely low signal quality (SS ≤ 4), however, the HDR processing algorithm failed to compensate for the low signal quality effect on the RNFL thickness measurement. In our experience, it is a rare case where the best achievable SS is less than 5. In cases of extremely

low signal quality, repeated scanning usually results in a scan with higher SS. It is relatively common that the best SS is 5, especially with older patients with some ocular pathology, which is less than the manufacturer's recommended acceptable SS cutoff of 6. One example is a case with one scan with a SS = 5 and another scan with a SS = 10 on the same eye that showed more than a 10 μm difference in RNFL thickness (Device slope 2.6 μm × 5 (SS difference)). However, with HDR processing, the expected difference reduces to -0.05 μm (HDR slope -0.01 μm), which is negligible. Therefore, images with a SS of 5 may become acceptable for clinical assessment.

HDR processing also enhanced the visualization of retinal layers and decreased the frequency of segmentation errors that are common in poor SS images. The HDR processing method divided data from a single OCT image into three virtual channels based on the histogram distribution, mimicking the low, medium, and high exposure images used in the traditional HDR technology in photography. By stretching each channel to the full 8-bit data dynamic range, optimized tissue visualization can be obtained.

Strictly speaking, the presented HDR processing technique is not a pure HDR technique because three different "virtual exposure" images are created using the same original OCT data. However, by expanding low, medium, and high signal channels, detailed tissue information was enhanced and became visible, which generated an effect similar to the outcome of the actual HDR technique.

For objective image quality evaluation, images showed significantly lower SNR but higher CNR than the original images after HDR processing. The decrease of SNR is due to boosting the speckle noise along with the meaningful signal by expanding the low signal channel. On the other hand, the improvement of CNR agrees with the subjective assessment that HDR processing enhances the visibility of fine details of the retinal tissues. Observers tended to

look at the signal quality on the homogeneous parts of the retina and compare it against the background, which is the same as how CNR was calculated (Eq. 3-4 and Eq. 3-5). The CNR measures the differences between the signal quality of the homogeneous parts of the retina and the noise signal. Both the retinal signal quality and the noise signal quality were raised after HDR processing, but the effects on the meaningful signal outperformed the adverse effect on the boosted noise. This indicates that for clinically acceptable high signal quality images, the proposed HDR processing method has the ability to improve image quality and enhance the visibility of the fine details of the retinal tissues, which may help better clinical diagnoses and image reading.

Though the HDR processing increased the noise signal and thus resulted in the reduction of SNR, the HDR processing still eliminated some segmentation failures. As with the effects of subjective assessment, a similar positive effect was observed on segmentation performance even with the boosted noise. Combined with its compensation effect on RNFL thickness measurement variability and possibly with a noise reduction method, the HDR processing may provide better RNFL thickness measurement reproducibility, and improve the accuracy of longitudinal clinical assessment on disease management.

In conclusion, the novel application of a standard signal enhancement method based on HDR imaging concept successfully restored OCT signal and image quality for both TD- and SD-OCT images. We further confirmed that the RNFL thickness measurement differences caused by image quality variation were significantly reduced to the expected measurement variability.

3.4   HISTOGRAM MATCHING

Another image processing method used to compensate for the measurement variability related to image quality variation is presented here. Histogram matching (HM) is an image processing technique to calibrate the differences in intensity contrast when capturing images with different cameras, image acquisition equipment, settings, and different light sources.[100, 101] By shaping an input image histogram to a reference histogram, HM is able to compensate for the differences in intensity and image contrast, and even enhances the image quality. HM technique has been widely used as a pre-processing step in cellular imaging and many medical imaging modalities, such as positron emission tomography (PET), single photon emission computed tomography (SPECT), and MRI, to correct the difference in background intensity and improve the registration and analysis variability.[102]

We hypothesized that the application of HM in OCT image data will enhance the image quality of the images with lower signal quality, and therefore be able to reduce the RNFL thickness measurement variability related to image quality variations. The purpose of this section was to minimize the influence of signal quality related OCT RNFL thickness measurement variability using a novel OCT image processing method utilizing HM technique.

3.4.1   Subjects and Image Acquisition

This was an observational cross-sectional study. Healthy volunteers were recruited at the University of Pittsburgh Medical Center Eye Center. The circumpapillary region from all eyes was scanned using Cirrus HD-OCT (software version 6.5) with the Optic Disc Cube 200×200 scan pattern as described in Section 2.2.1. A series of OCT scans (at least 10 scans) were

acquired from each eye at the same visit to achieve a wide range of SS (a metric of OCT image quality is provided by the device manufacturer that ranges from 1 to 10) by intentionally defocusing and changing the refraction settings. The image inclusion criteria were mentioned in Section 2.2. Raw OCT image data files were exported to a standalone computer for further processing and analysis.

3.4.2 Histogram Matching Algorithm Development

The HM processing is divided into three steps: Circular B-scan resampling, then speckle noise reduction, followed by reference histogram construction or HM processing, depending on the SS. The overall flow of HM processing is presented in Figure 11. The details of each step are described below.

Circular B-Scan Resampling

For each 3D cube image data, the geometric center of the ONH was automatically determined based on the manually delineated disc margin. 512 equally spaced A-scans were sampled along the 3.4 mm diameter circle to generate a virtual circular Cirrus B-scan.

Speckle Noise Reduction

The same speckle noise reduction method as described in Section 3.2 was applied to the resampled circular B-scan here to reduce the speckle noise. In brief, the speckle noise reduction method was a localized high amplitude signal removal method based on a selective smoothing method where only spiky OCT signals are suppressed to the level of its heavily smoothed counterpart. In this way, high frequency components with relatively high amplitude, considered

as speckle noise, were removed, while high frequency components with low amplitude, recognized as retinal tissue signals, were preserved so that more details of the tissue structures remained in the OCT signals after noise reduction.[103]
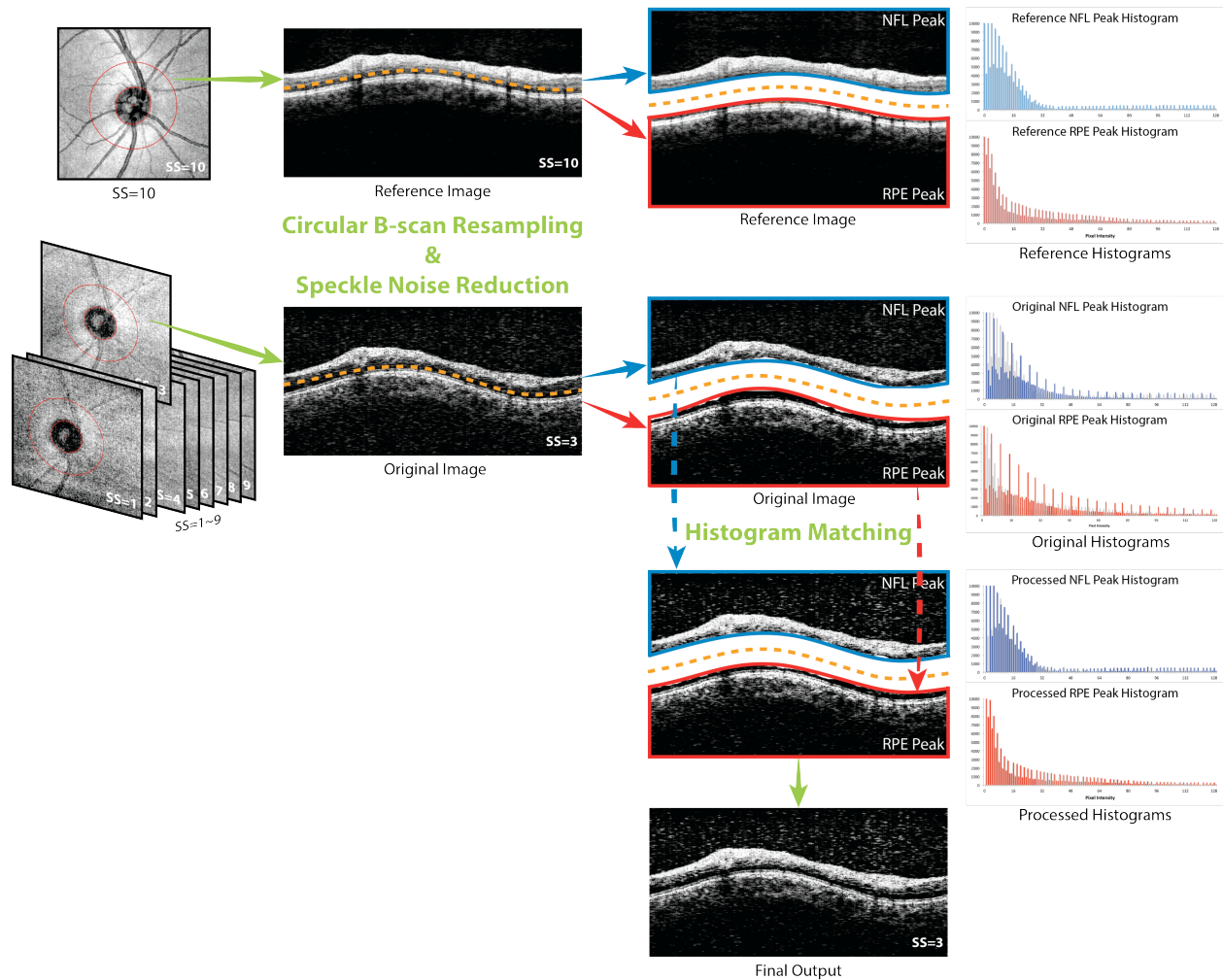
Figure 11. Flow chart of the HM method. Circular B-scan resampling, then the speckle noise reduction, followed by reference histogram construction or HM processing depending on the SS.

Reference Histogram Construction (applied only for the image with the highest SS for each eye)

After circular B-scan resampling and speckle noise reduction, the histogram of the resampled image with the highest SS of each image series was set as the reference histogram. Images with lower SS were processed with the HM method (described below) so that their histograms had the same shape as the reference histogram.

To take the clinic reality into account, where for some elderly or diseased eye, images with a good SS (SS>6) cannot be achieved, reference histograms were constructed in two ways (Figure 12):

I.   **Individual reference histogram.**

With the speculation that histogram statistics differ among retinal tissues, we partitioned the circular B-scan image into top and bottom halves along the valley, where the OCT signal amplitude was the lowest between the outer plexiform layer (OPL) and the external limiting membrane (ELM) (the yellow dash curve in Figure 12-I). The valley was automatically detected by our segmentation software, which is based on the algorithm described elsewhere.[85] The top half contained the vitreous body, RNFL, ganglion cell layer (GCL), IPL, and OPL, and therefore is also called the NFL peak. The bottom half included the outer nuclear layer, inner and outer segment (IS/OS), RPE, and the region below the RPE and therefore was also noted as the RPE peak (Figure 12-I). The total pixel numbers of the top and bottom halves were matched to half of the entire circular B-scan by padding or cropping signals from the region in the vitreous or below the RPE so that the proportion of actual retinal signal was consistent across all the subjects.

For each image series, the histograms of the top and bottom halves of a resampled image with the highest SS were set as the reference histograms for the top and bottom halves respectively. In other words, each image series had its own top and bottom reference histograms.

57

II.   Group reference histogram.

The mean histogram patterns of all the reference histograms were generated for the top and bottom halves separately. The mean outcome histograms were used as the group reference histograms for all the images.

Histogram Matching (applied to the rest of the images)

HM has been known as an image processing technique where a series of histogram equalization steps is used to match the statistical information, or histogram shape, of two images. For each image series, all images except for the one with the highest SS were processed with HM. All subjected images were pre-processed and partitioned in the same way as the reference histogram.

To begin with, the percentile information on the histograms at each intensity was calculated by Eq. 3-8, as shown in Figure 13.[104]

$$P[i] = \frac{\sum_{x=0}^{i} n_x}{N}, 0 \leq i \leq 255 \qquad \text{(Eq. 3-8),}$$

where $P[i]$ indicates the percentile information at intensity i, while n_x shows the number of pixel having intensity x, and N is the total pixel number within the region. For Cirrus data, the full intensity dynamic range is from 0 to 255.
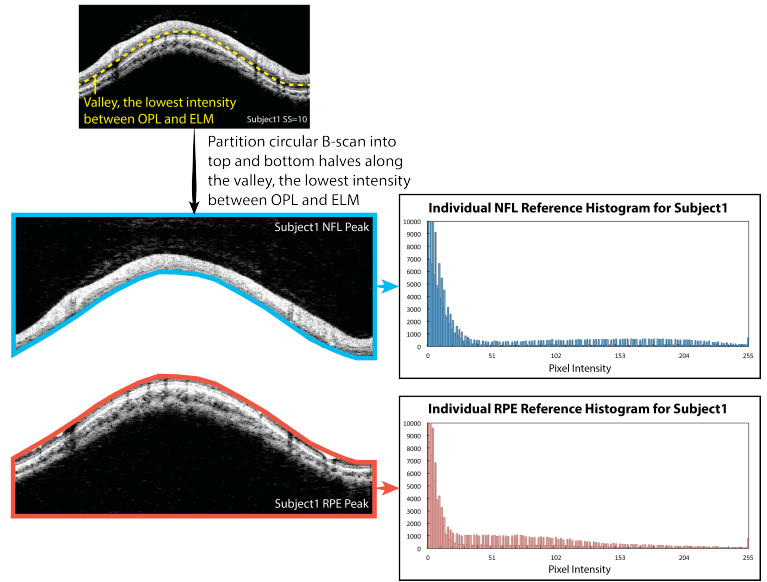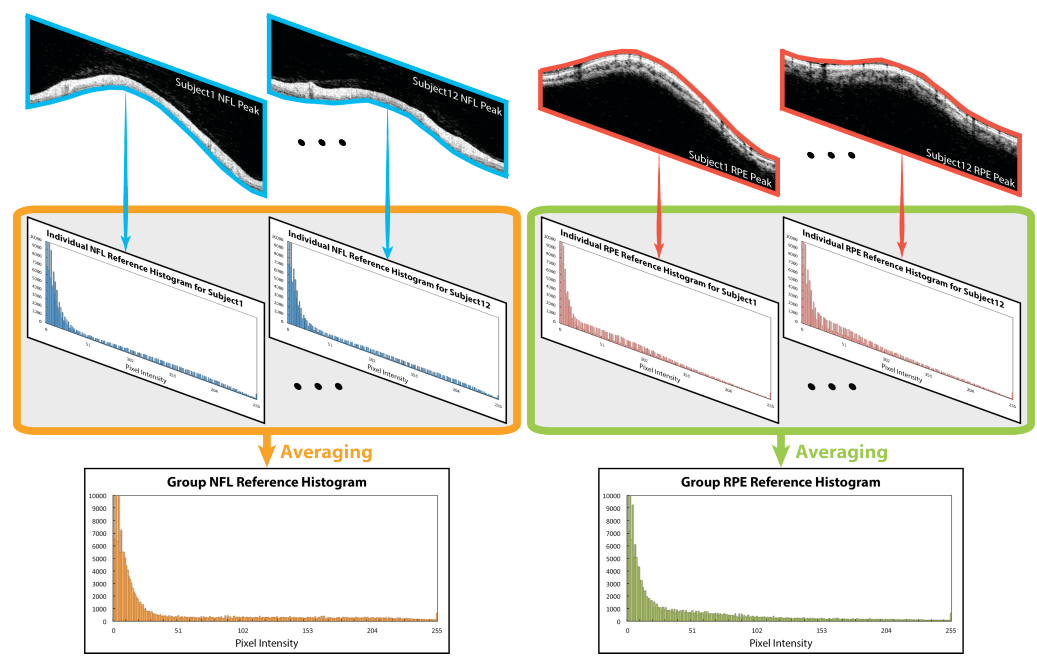
Figure 12. The schematic figures of how the reference histograms were constructed. (I) Individual reference histogram (partitioned into top and bottom halves, or the NFL and RPE peaks), and (II) Group reference histogram. The blue and red borders surrounding the circular B-scan indicate the region where the reference histograms were calculated. The range of the vertical axis is adjusted for better visualization of the histogram.
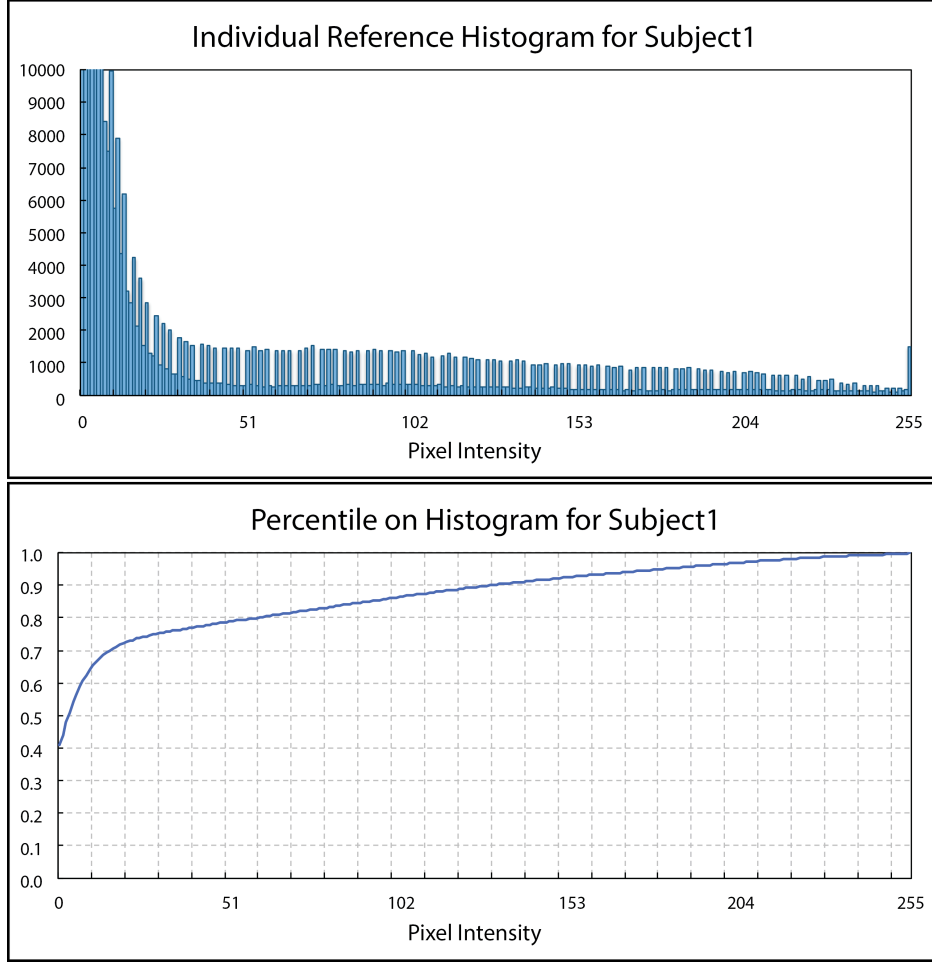
Figure 13. The reference histogram and corresponding percentile information.

Based on the percentile information, a mapping matrix was used to convert the shape of the input histogram to the shape of the reference histogram was then generated by matching or minimizing the distance in percentile at each intensity on the histograms between reference and input histograms, as shown in Eq. 3-9.

$$T[i] = j, \text{ if } \left| P_{in}[i] - P_{ref}[j] \right| = \min_{k} \left| P_{in}[i] - P_{ref}[k] \right| \qquad \text{(Eq. 3-9)},$$

where T[ ] is the resulted mapping matrix, $P_{in}[i]$ and $P_{ref}[j]$ are the percentile information on the input and reference histograms at intensity i and j, respectively. For each intensity i in the input histogram, we found a corresponding intensity j in the reference histogram, so that the

percentile $P_{in}[i]$ and $P_{ref}[j]$ had minimal difference. All the sample points in the input image data with intensity i were then mapped to intensity j to generate the output image data.

Conventional HM procedure views the sampling points with the same intensity as a group and thus cannot distinguish pixels with the same intensity. This generates the approximation errors due to quantization and rounding-off, which can be observed as the spiky shape of the outcome histogram as Figure 14 shows.[101, 105, 106] To solve the approximation errors and enable separating pixels with the same intensity, we introduced a sub-feature besides the intensity to each sampling point.[101, 105, 106] The sub-feature virtually made the histogram bin finer than the minimal intensity unit and allowed us to have more flexibility to model the histogram shape.[101, 105, 106]

The sub-feature we added here was the mean intensity of a pixel's 3×3 neighbors. With this sub-feature, each sampling point had a new value, I_{new}, as presented in Eq. 3-10:

$$I_{new} = I_{ori} \times 256 + I_{mean} \qquad \text{(Eq. 3-10)},$$

where 256 is the full intensity dynamic range for Cirrus data, while I_{ori} and I_{mean} stand for the original intensity of the sampling point and the mean intensity of its 3×3 neighbors.

HM and mapping matrices were performed and generated based on the new intensity and corresponding histogram. The effects of the individual and group reference histograms were tested separately.
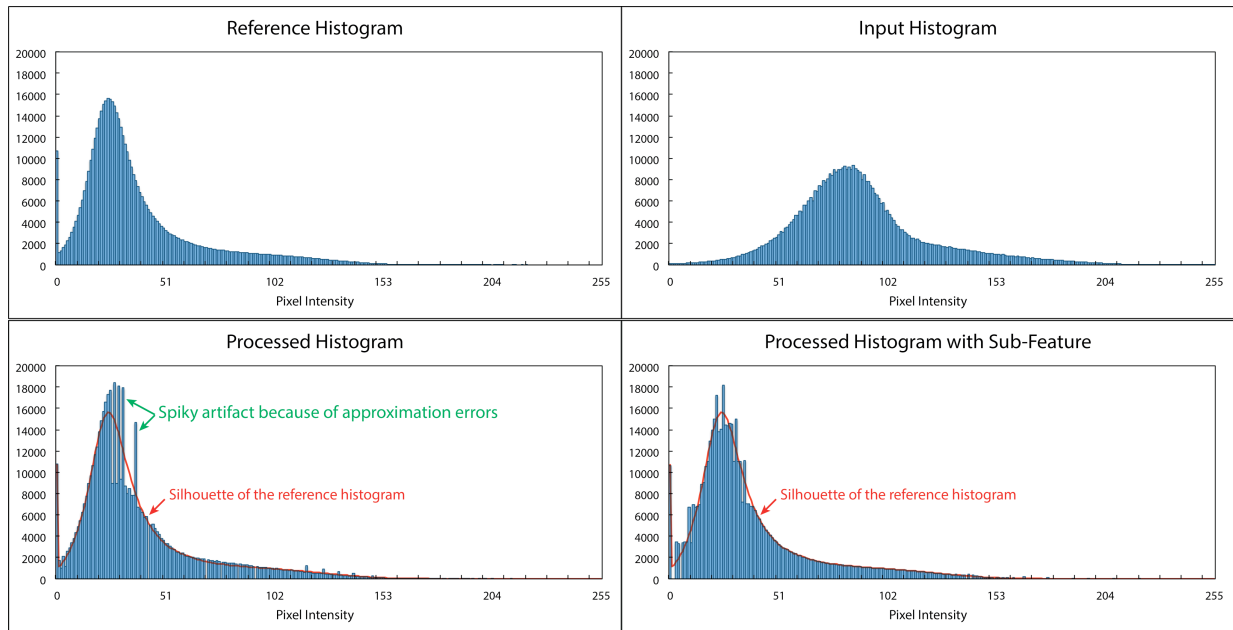
Figure 14. HM with sub-feature reduces the quantization and rounding-off errors. Top left and right: the reference and the input histogram. Bottom left and right: the histogram using conventional HM and using HM with sub-feature. The quantization and rounding-off errors (spiky artifacts as indicated by the green arrows in the bottom left figure) can be reduced with the sub-feature, which enables us to separate pixels with the same intensity but in different retinal layers, as shown in the bottom right figure. The red curve in the bottom left and bottom right figures present the shape of the reference histogram.

3.4.3 Histogram Matching Performance Assessment

RNFL Thickness Measurements

The proposed HM processing was tested by comparing the circumpapillary RNFL thicknesses before and after HM processing. The RNFL thicknesses were measured using our custom segmentation software (the same algorithm used in Section 3.3.2) both before and after

62

HM (Original and HM measurements). The thickness measurements reported from the Cirrus machine (Device measurement) were also collected. The relationships between Device, Original, and HM measurements and SS were further investigated.

Statistical Analysis

Wilcoxon tests were used to compare the total measurement variability (the maximum measurement variability with SS range from 1 to 10) among Device, Original, and HM measurements. Broken stick non-linear mixed effects models were applied to analyze the relationship between RNFL thickness for Device, Original, and HM measurements. In addition, the coefficient of variation (CoV) of the Device measurements within the manufacturer recommended acceptable SS range (SS from 6 to 10) was calculated on each case. The lowest SS that achieved similar Device CoV on Original and HM measurements was detected for each case, in order to see if the acceptable SS range could be extended with the present method.

3.4.4 Results

Twelve right eyes from 12 healthy volunteers (4 males and 8 females) were recruited at the University of Pittsburgh Medical Center Eye Center. The average age was 31.7 ± 11.1 years with visual field average MD of -0.21 ± 1.46 dB.

For the relationships between RNFL thickness and SS, two segments of linear relationship were detected on Device, Original, and HM measurements. Table 5 summarizes the slopes and intercepts of both segments, and the breaking points of all the measurements. An example is presented in Figure 15. The slope in the first segment (β_1) presents the linear relationship in the lower SS range (range from 1 to the breaking point), while the slope in the

second segment (β_2) presents the linear relationship in the higher SS range (range from the breaking point to 10). The Original measurements had statistically significantly larger slope in the second segment than the Device measurements (4.89 vs 1.72 µm/SS), indicating that Original measurements were more sensitive to SS. This strong correlation of the Original measurements was successfully reduced after HM, regardless of using individual or group reference histograms (1.17 and 1.06 µm/SS).

Table 5. Statistical analysis summary. The slopes of first and second segment (β_1 and β_2), intercepts at SS=0 (α_1), intercepts at the breaking point (α_2), and the breaking point for the relationships between RNFL thickness and SS of Device, Original, and HM measurements. (95% CI is shown in the parentheses).

		α_1	β_1	α_2	β_2	**Breaking point**
Device		-28.91	42.57	88.16	1.72	2.75 (2.49, 3.00)
Original		35.86	11.77	105.77	4.89	5.94 (4.93, 6.95)
Histogram Matching	**Individual Reference**	71.73	14.62	119.39	1.17	3.26 (2.70, 3.82)
	Group Reference	70.84	14.89	120.27	1.06	3.32 (2.81, 3.83)
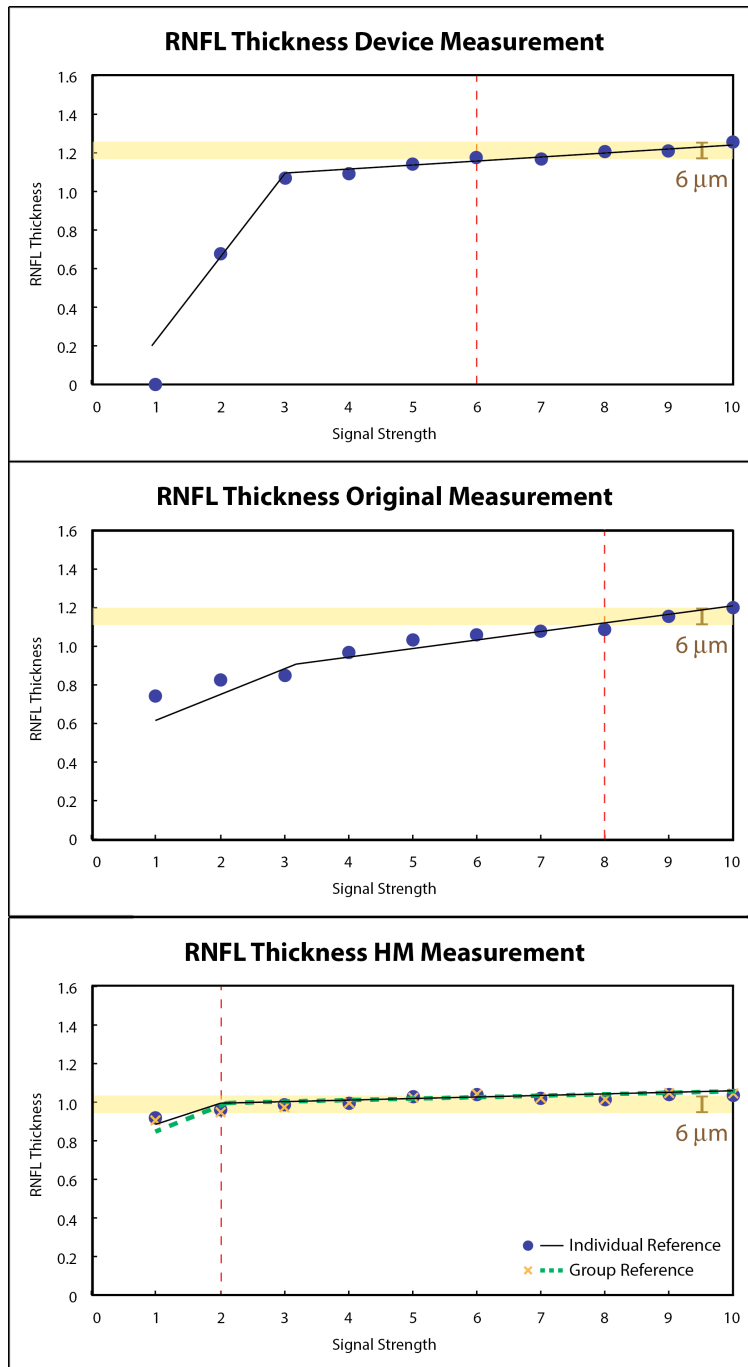
Figure 15. Relationships between RNFL thickness and SS. RNFL thicknesses from one subject were plotted against SS with two segments of the linear relationships on the Device (top), Original (middle), and the HM measurements (bottom). The Original measurements show larger measurement variability. While with HM, a lower SS (SS=2) was achieved to maintain the same maximum absolute difference (yellow band) within the manufacturer recommended acceptable SS range on the device (top, SS=6), as indicated by the red vertical dash line.

65

For the breaking point, the Original measurements had statistically significantly higher breaking points than the Device measurements (5.94 vs 2.75 SS), while it became similar after applying HM (3.26 with individual reference and 3.32 with group reference), suggesting that HM stabilized the measurement variability in a wider SS range.

No statistically significant differences in total measurement variability (SS range 1 to 10) were found between Device and Original measurements (p=0.86, Wilcoxon tests) (Table 6). However, HM measurements showed statistically significantly smaller total measurement variability than the Device measurements on both individual and group references (33.4 and 33.5 μm, both p<0.038, respectively, Wilcoxon tests).

The mean Device CoV for the recommended acceptable SS range (SS 6 to 10) was 0.025. The original measurements reached the same CoV with a SS ranged from 8.7 to 10. HM measurements showed the same CoV with a wider SS range (SS 3.0 to 10 and 2.2 to 10, for individual and group reference histograms respectively).

Table 6. Summary of the measurement variability and minimum acceptable SS range. In the measurement variability column, 95% CIs are shown in the parentheses. *: Significantly difference between the method and Device measurements

		Measurement Variability (μm) SS: 1-10	Minimum Acceptable SS
Device (Reference)		64.83 (50.94, 78.73)	6
Original		69.11 (55.21, 83.01)	8.7
Histogram Matching	**Individual Reference**	33.39 (19.49, 47.29)*	3.0
	Group Reference	33.56 (19.66, 47.46)*	2.2

3.4.5 Discussion

A novel HM based OCT image enhancement method was developed. The proposed method successfully enhanced OCT images with lower image quality and reduced the RNFL thickness measurement variability related to image quality variation. With the enhanced image quality and reduced thickness measurement variability, HM further extended the acceptable signal quality range, which would broaden the application of OCT to elder or diseased subjects who tend to have lower best-achievable image quality.

In the HM processing, a sub-feature was added to separate pixels with the same intensity but different characteristics, which should not be classified into the same group. Any feature that is able to make two pixels that have the same intensity but carry different information distinguishable can be used as a sub-feature, such as the location of the pixel in the retina, mean neighbor intensity (with various neighbor regions, such as 3×3, 5×5, or 1×3 neighbor pixels within the same A-scan), and the variance among neighbor pixels. The more distinguishable the sub-feature is, the finer HM we can achieve. However, simply applying the most distinct sub-feature, for example the axial position of the pixel, did not improve the results because the axial location of the retina varies within a frame. The sub-feature needs to be able to separate pixels with different characteristics, but keep the similar pixel in the same group. On the other hand, the variance among neighbor pixels has less separation power. Therefore we chose to use the mean intensity of a pixel's 3×3 neighbor as the sub-feature. By adding mean 3×3 neighbor intensity as the sub-feature, the contextual information (surrounding tissue information) can be included in the HM process to improve the outcome.

Separating the retina into two parts and applying individual HM on each part improved the image enhancement performance. We observed that the OCT signal from RPE peak was

generally stronger than the signal from the NFL peak along all SS variation. In other words, the signal from the NFL peak degraded more than the signal from RPE peak as SS decreased. Therefore, when simply applying HM to the entire image, the signal from NFL peak remained weak even after enhancement. This affected the segmentation performance as the NFL peak is supposed to have a relatively high reflectivity in principle. When partitioning the circular B-scan into the top and bottom halves, the NFL signal was boosted to a similar level of RPE, which made the segmentation performance more reliable and led to reduced measurement variability.

No statistically significant differences were detected in both the slope of the second segment and the breaking point when using individual and group reference histograms. This indicates that the specific reference histogram for each image series can be replaced by a group reference histogram and similar performance can be achieved. However, there is a limitation. The group reference histogram works based on the assumption that the disease does not change the histogram characteristics of the scan. If the histogram characteristics are different from the group reference histogram, then the outcomes may not reflect the actual disease status. In that case, an individual reference histogram or a separate group reference histogram with similar histogram characteristics (pathology) is required. Further investigation for the validation of such references in warranted.

With HM processing, similar RNFL thickness measurements were generated across a wider SS range, suggesting less measurement variability expected even with lower SS image than the current manufacturer's recommendation. In addition, the measurement variability was also reduced across the recommended SS range (SS 6 to 10). The reduced measurement variability may help reduce the false positive reading due to acceptable but low SS.

One major limitation of this study was our use of custom segmentation algorithm to test the proposed HM method because the device software does not accept modified OCT image files for processing. Strictly speaking, the observed improvement is limited to the custom algorithm. However, the overall trend in correlation between segmented thickness measurements and image quality is observed regardless of the differences in algorithm approach. Therefore, it is reasonable to speculate that the HM method may expand the acceptable SS range without affecting the OCT measurement variability.

In conclusion, the proposed HM method successfully enhanced OCT images and extended the acceptable SS range on OCT images. With the potential to achieve a wider acceptable SS range, HM would qualify more OCT images with relatively low SS for clinical assessment, and further broaden the OCT application to a wider range of subjects.

4.0 QUANTITATIVE ASSESSMENT OF THE SIGNAL NORMALIZATION

In Chapter 3, a signal normalization method was developed. Multiple image processing methods were integrated in a stepwise fashion to overcome various factors that cause OCT signal characteristics variability, such as sampling density normalization, speckle noise reduction, amplitude normalization, and image quality compensation. To test the ability of the signal normalization in minimizing the discrepancies in signal characteristics and data measurement differences among SD-OCT devices, quantitative assessments were conducted in two ways: engineering validation and clinical validation. In the engineering validation, the absolute differences between individual A-scans were measured to investigate the effects of the normalization method. In the clinical validation, the systematic differences in RNFL thickness among SD-OCT devices were assessed before and after signal normalization to assess the effect of signal normalization on actual clinical measurements.

4.1 ENGINEERING VALIDATION OF THE SIGNAL NORMALIZATION

In the engineering validation, we measured the absolute differences between individual A-scans from two SD-OCT devices, namely Cirrus and RTVue, to investigate the effects of the developed signal normalization on reducing the signal characteristics variability.

4.1.1 Methods

A total of 14 healthy and 7 glaucoma subjects volunteered to participate in this prospective cross-sectional study. One eye from each subject was randomly selected and used in the study. For the diseased eyes, a variety of glaucoma damage was included. The diagnosis of glaucoma was clinically defined based on the presence of visual field analysis and typical glaucomatous structural changes. The inclusion criteria were the same as described in Section 2.0 and Section 2.1.

For the instruments and image acquisition, the macular region was imaged at the same visit using two commercially available SD-OCT devices with equivalent 3D cube scan patterns: Cirrus HD-OCT (software version 5.1) and RTVue (software version 6.1). Details of the scan protocols and image inclusion criteria can be found in Section 2.2.

4.1.1.1 Signal Normalization Processing

The experiment methods were divided into three stages: image registration and sampling, signal normalization, and A-scan profile comparison. As mentioned in Chapter 3, Cirrus OCT data format was used as the normalization reference data format, so the RTVue OCT data format was converted to a Cirrus-equivalent OCT data format.

Image Registration and Sampling

In order to sample the A-scan profile from the same location from different OCT data, the foveola position was manually selected on both the Cirrus and RTVue cube data by looking for the largest separation between the junction of the IS/OS of the photoreceptors and RPE as appearing on the horizontal and vertical cross-sectional B-scans (Figure 16). The selected

foveola position was then used as the center for both registration and sampling. RTVue data were then translated and rotated to match the blood vessel position subjectively on the Cirrus *en face* image by finding the translation vector and rotation degree which minimized the absolute differences in pixel value between Cirrus and RTVue *en face* images. After registration, single A-scans were sampled 1.8 mm from the foveola in the temporal, superior, nasal, and inferior quadrants from Cirrus and RTVue data and were saved for further processing and analysis. The sampling position of each A-scan pair in Cirrus and RTVue was subjectively evaluated according to the relative position of major blood vessels. If the A-scan pairs were sampled from different positions (absolute distance larger than the width of a major retinal blood vessel) in Cirrus and RTVue, the A-scan pairs were excluded. Figure 17 shows the flow of the image registration and sampling.
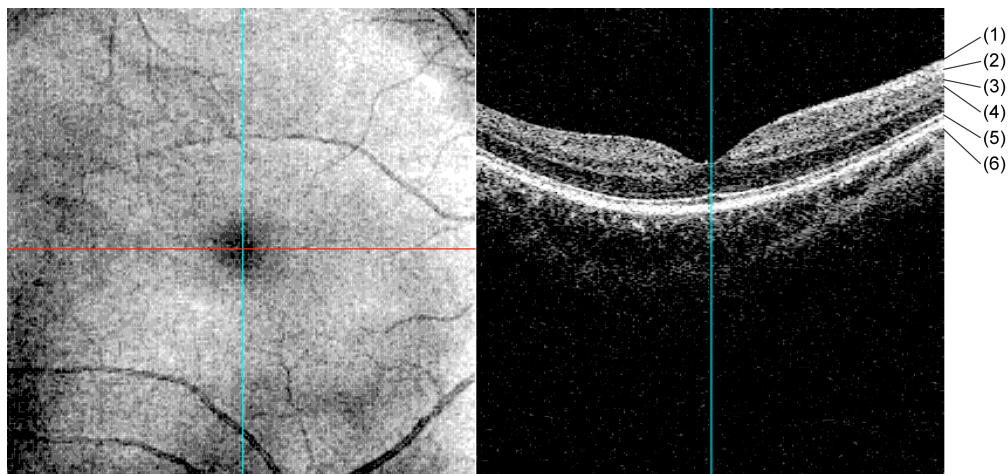
Figure 16. Demonstration of how to determine the foveola position. Foveola position was selected manually by looking for the largest separation between the junction of the IS/OS (5) and RPE (6). The red line on the *en face* image (left) indicates where the cross-sectional image on the right was sampled. The vertical cyan line indicates the location of the largest separation between the IS/OS and RPE. The intersection of the red and cyan lines on the *en face* image is the selected foveola position. (1) ILM, (2) NFL, (3) GCL, (4) Inner plexiform layer (IPL), (5) IS/OS, and (6) RPE.
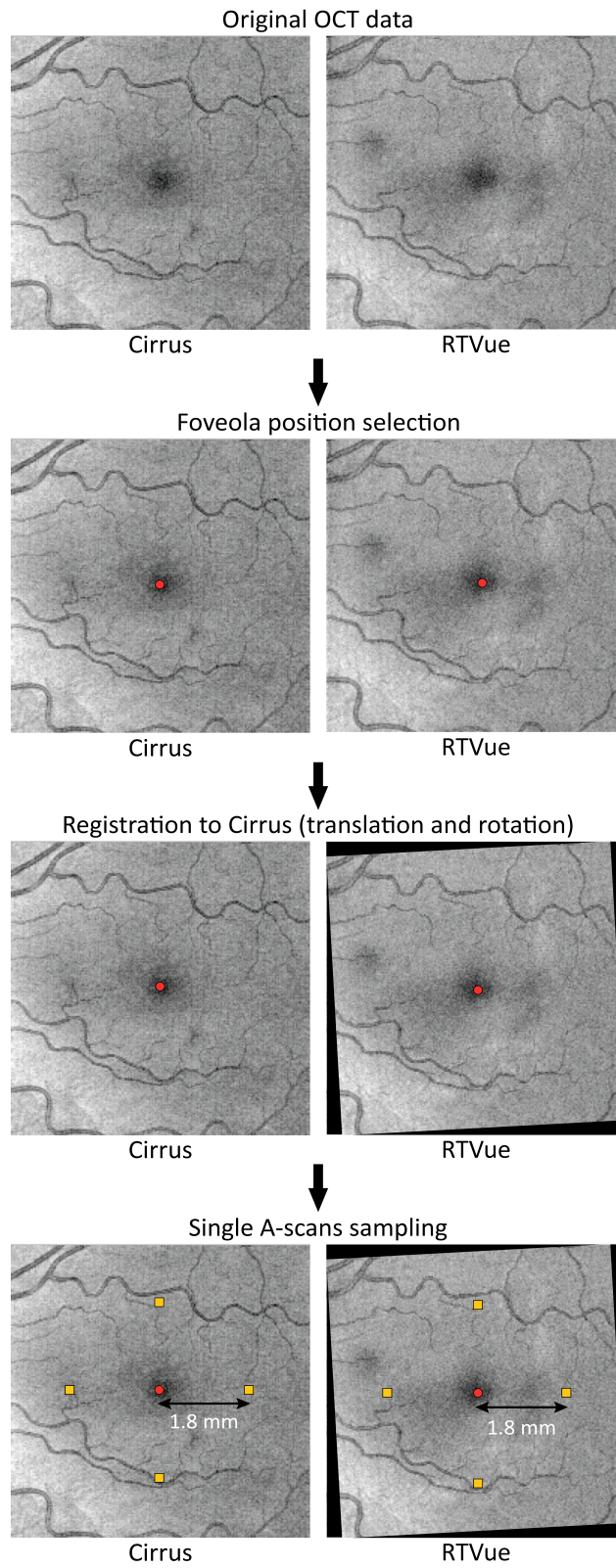
Figure 17. Image registration and A-scan sampling process.

Signal Normalization

The signal normalization consisted of three processing steps: z-scaling and sampling density normalization, speckle noise reduction, and amplitude normalization. Here the first step in speckle noise reduction was separated and used as individual amplitude normalization. The experiment was divided into two phases. In phase I, the effects of reducing the differences between Cirrus and RTVue in each processing step were assessed individually. In phase II, all three processing steps were combined together using different cutoffs, and the final results were assessed.

Phase I

The goal in phase I was to assess and optimize the ability to reduce the difference between Cirrus and RTVue OCT signals in each processing step. The testing started with z-scaling and sampling density normalization; secondly, speckle noise reduction and amplitude normalization were applied separately and the effects with various cutoffs were assessed.

Z-scaling and Sampling Density Normalization. The details of the normalization method were described in Section 3.1. After interpolation, RTVue was further linearly compressed from 12-bit to 8-bit data format. In this way, RTVue and Cirrus data would be at a comparable intensity level. The z-scaling and sampling density normalized 8-bit RTVue data were then used as the baseline RTVue data, and further processing was performed on them.

Speckle Noise Reduction. After z-scaling and sampling density normalization, the previously described speckle noise reduction method was applied to reduce the speckle noise. The details of the method were similar to Section 3.2.1 except that instead of using 66th percentile on the histogram as a fixed low cutoff (noise level) in Step 3 of speckle noise

reduction, various low cutoffs were used in this step to test and optimize the effect of speckle noise reduction processing. To test the effects of the various cutoff thresholds in the mask amplitude matching (Step 3), the first, 33rd, 50th, and 66th percentiles on the histogram were used as the low cutoff. The high cutoff was always the 99th percentile on the histogram.

Amplitude Normalization. Amplitude normalization was designed to normalize the noise level between Cirrus and RTVue data. Various cutoffs were used for amplitude normalization to find the optimal settings: 1) A histogram-based amplitude normalization method as described in Section 3.2.1 *Step 1* was used here to linearly map the signals between the low cutoffs (the 1st, 33rd, 50th, and 66th percentiles on the histogram of the frame where the sampled A-scan was located) and high cutoff (the 99th percentile on the histogram of the frame where the sampled A-scan was located) to the full 8-bit gray scale level on OCT data for both Cirrus and RTVue, or 2) Data range matching: matching the data range of two A-scan profiles by linearly mapping the minimal and maximal intensity of the frame where the sampled A-scan was located to the full 8-bit gray scale for both Cirrus and RTVue data.

Phase II

In phase II, the three processing steps were combined and the outcomes of the combination of the three processing steps were evaluated. We combined the processing in the order that z-scaling and sampling density normalization came first, followed by speckle noise reduction, and then amplitude normalization with the cutoffs showing the best results in phase I.

A-scan Profile Comparison

After each step, the residual between Cirrus and RTVue A-scan pairs sampled from four quadrants were measured. A-scan pairs sampled from the same quadrant were aligned to the ILM. Then, the mean absolute difference in amplitude at each sampling point within the eligible measurement range was calculated, where the eligible measurement range means that within this range, all the A-scan pairs were able to find corresponding Cirrus and RTVue data. Since original Cirrus and Z-scaled RTVue data had different data ranges as compared to normalized Cirrus and RTVue data, the mean absolute difference in amplitude was normalized to the percentage of the maximal data range from the two A-scans in order to compensate for the data range inconsistency, as shown in Eq. 4-1. The absolute difference in amplitude between Cirrus and RTVue data is presented as the shaded region between Cirrus and RTVue A-scan profiles (Figure 18). The mean absolute difference in percentage between Cirrus and Z-scaled RTVue data was used as the baseline difference between the two OCT devices. Mean absolute difference in percentage between two Cirrus scans, acquired from the same eye at the same visit sampled at the same location, was computed and used as the reference for similarity assessment. The same eligible measurement range was applied to calculate the mean absolute difference in percentage between two Cirrus scans.

$$\text{Mean Absolute Difference (\%)} = \frac{\sum_i \left(\frac{|Cirrus_i - RTVue_i|}{Max\ Data\ Range} \right)}{Eligible\ Measurement\ Range} \times 100\% \qquad \text{(Eq. 4-1)}.$$
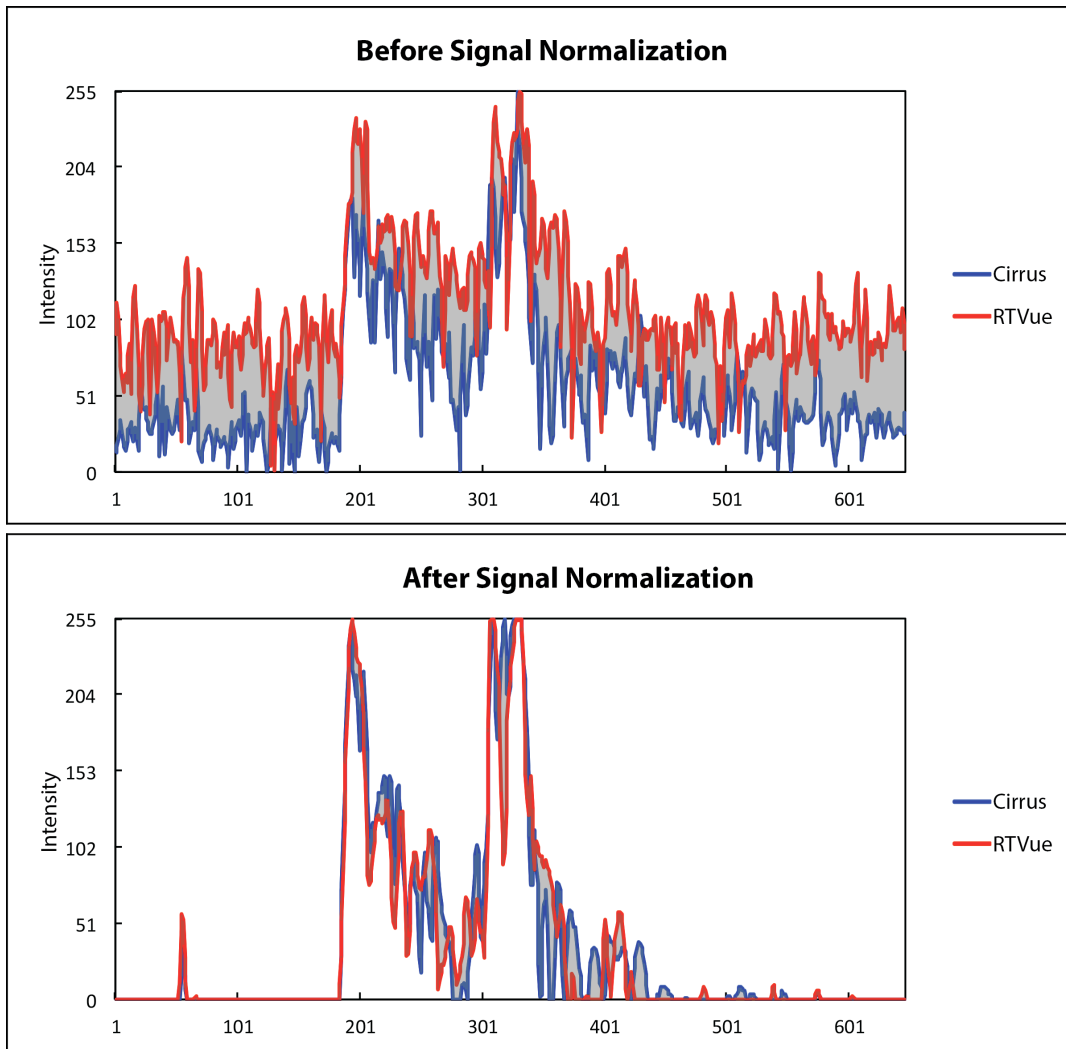
Figure 18. Absolute difference between A-scan profiles before and after signal normalization. Cirrus (blue line) and RTVue (red line) A-scan profiles as recorded within the eligible measurement range. The shaded area between Cirrus and RTVue is the residual between the two A-scan profiles, and is used as a quantitative analysis parameter.

4.1.1.2 Statistical Analysis

Paired t-tests were used to analyze the overall and quadrant absolute differences between original Cirrus and Z-scaled RTVue data, between normalized Cirrus and normalized RTVue data, and between two original Cirrus scans. $P < 0.05$ was considered as statistically significant.

4.1.2 Results

Subject demographics are presented in Table 7. Fourteen healthy and 7 glaucoma subjects were enrolled in this study. Healthy eyes were younger than glaucomatous eyes (41.9 ± 16.9 vs 65.2 ± 5.5 years, p=0.0023, t-test).

Table 7. Subject demographics.

	Healthy (n=14)	Glaucoma (n=7)
Male / female	4:10	1:6
OD / OS	10:4	4:3
Age (years)	41.9 ± 16.9	65.2 ± 5.5
MD	0.6 ± 0.6	-2.0 ± 2.0
Total retinal thickness	303.8 ± 12.5	297.8 ± 8.2

Table 8 summarizes the mean absolute difference in amplitude in the percentage of A-scan profiles between the original Cirrus and Z-scaled RTVue data (baseline residual), and between Cirrus and RTVue data after speckle noise reduction with various cutoff settings. The overall residual was statistically significantly reduced after speckle noise reduction with the settings using the 33rd percentile on the histogram as the low cutoff and the 99th percentile as the high cutoff (p=0.0031, paired t-test), but not when using the 50th and 66th percentile on the histogram as the low cutoff. When using the first percentile on the histogram as the low cutoff, the residual was significantly increased (p<0.0001, paired t-test). Among all the settings, using the 33rd percentile on the histogram as the low cutoff statistically significantly outperformed the settings with the rest low cutoffs (p<0.0001), and generated the largest reduction in the residual

in amplitude between Cirrus and RTVue. For quadrant analysis, only the superior quadrant with the 33rd percentile on the histogram as the low cutoff showed a statistically significant reduction compared to the baseline residual. For the 50th and 66th percentile on the histogram as the low cutoffs, there was no significant difference in the residual compared to baseline in all four quadrants. When using the first percentile on the histogram as the low cutoff, the residuals significantly increased in all the quadrants. A similar trend in the overall results was found in the quadrants analysis that using the 33rd percentile on the histogram as the low cutoff statistically significantly reduced the baseline residual the most when compared to other settings ($p < 0.0048$).

Table 8. Mean absolute difference in amplitude between Cirrus and RTVue A-scan profiles I. Speckle noise reduction was used as the next step following z-scaling and sampling density normalization. Absolute difference is in the percentage with the difference of baseline the residual from the residual after speckle noise reduction in parentheses. The minus sign indicates that there was a reduction in the residual between Cirrus and RTVue after speckle noise reduction.

	Overall (%)	Temporal (%)	Superior (%)	Nasal (%)	Inferior (%)
Baseline residual	12.7	12.8	12.4	13.0	12.6
First percentile	14.6 (1.9)	15.9 (3.1)	13.3 (0.9)	15.5 (2.5)	13.7 (1.1)
33rd percentile	12.4 (-0.3)	12.7 (-0.1)	12.1 (-0.2)	12.6 (-0.5)	12.3 (-0.3)
50th percentile	12.7 (-0.0)	13.0 (0.3)	12.4 (0.1)	12.8 (-0.2)	12.5 (-0.0)
66th percentile	12.8 (0.1)	13.1 (0.3)	12.6 (0.2)	12.9 (-0.1)	12.6 (-0.0)

Table 9 shows the mean absolute difference in amplitude in the percentage of A-scan profiles between original Cirrus and Z-scaled RTVue data (baseline residual) and between Cirrus and RTVue data with amplitude normalization as the second step after z-scaling and sampling

density normalization using various cutoff settings. The overall residual was statistically significantly reduced when using the 50th and 66th percentile on the histogram as the low cutoff and the 99th percentile on the histogram as the high cutoff (both p<0.0001, paired t-test). No significant difference in the residual before and after amplitude normalization was found when using the 33rd percentile on the histogram as the low cutoff. A statistically significant increase in the residual was detected when applying the first percentile on the histogram as the low cutoff and using the data range matching method (both p<0.0001, paired t-test). Among the settings which successfully reduced the residual, the method using histogram-based amplitude normalization with the 66th percentile as the low cutoff statistically significantly outperformed other settings and contributed to the largest amount of reduction compared to baseline residual. For the quadrant analysis, the same trends as the overall results were found in all four quadrants.

Table 9. Mean absolute difference in amplitude between Cirrus and RTVue A-scan profiles II. Amplitude normalization was used as the next step following z-scaling and sampling density normalization. Absolute difference is in percentage with the difference of baseline residual from the residual after amplitude normalization in parentheses. The minus sign indicates that there was a reduction in the residual between Cirrus and RTVue after speckle noise reduction.

	Overall (%)	Temporal (%)	Superior (%)	Nasal (%)	Inferior (%)
Baseline residual	12.7	12.8	12.4	13.0	12.6
First percentile	20.5 (7.8)	20.8 (8.0)	20.2 (7.9)	20.8 (7.8)	20.0 (7.4)
33rd percentile	12.6 (-0.0)	12.6 (-0.1)	12.4 (0.0)	12.9 (-0.1)	12.5 (-0.1)
50th percentile	10.7 (-2.0)	10.6 (-2.2)	10.5 (-1.9)	11.0 (-2.0)	10.6 (-2.0)
66th percentile	8.7 (-4.0)	8.6 (-4.2)	8.5 (-3.9)	9.1 (-3.9)	8.6 (-4.0)
Data range matching	19.6 (6.9)	20.3 (7.5)	18.9 (6.5)	19.8 (6.8)	19.4 (6.8)

In phase II, for the combined signal normalization method, the three processing steps were combined in the following order: z-scaling and sampling density normalization, followed by speckle noise reduction, and at the end of the signal normalization we performed amplitude normalization. The settings which generated the best results of each processing step were used in the final combined method. For the speckle noise reduction step, the 33rd and 99th percentile on the histogram were used as the low and high cutoff. In amplitude normalization, the histogram-based amplitude normalization method with the 66th and 99th percentiles on the histogram as the low and high cutoffs was used to remove the noise level difference and match the different data range between Cirrus and RTVue.

An example of step-by-step signal normalization in phase II is presented in Figure 19. The first row in Figure 19 shows the original A-scan profile from Cirrus (blue) and RTVue (red). For display purposes, the dynamic range of RTVue data was linearly rescaled from 12-bit to 8-bit gray scale in Figure 19 (the first row). A-scan profiles from Cirrus and RTVue were aligned to ILM so the effect of signal normalization can be appreciated easier. As the first row in Figure 19 shows, the original A-scan profiles looked dissimilar and had different noise levels and sampling densities. The second row presents the results after z-scaling and sampling density normalization. After oversampling RTVue data in the axial direction, the sampling density of the two A-scan profiles became the same and the peaks in the A-scan profiles matched. However, there is still a noise level difference between Cirrus and Z-scaled RTVue data. The third row shows the results after speckle noise reduction. Compared to the second row, the high spiky peaks were removed, and the high frequency components with low intensity values were kept intact. The last row in Figure 19 shows the final results. After amplitude normalization, the noise level of the two A-scan profiles became the same and the A-scan profiles looked similar.
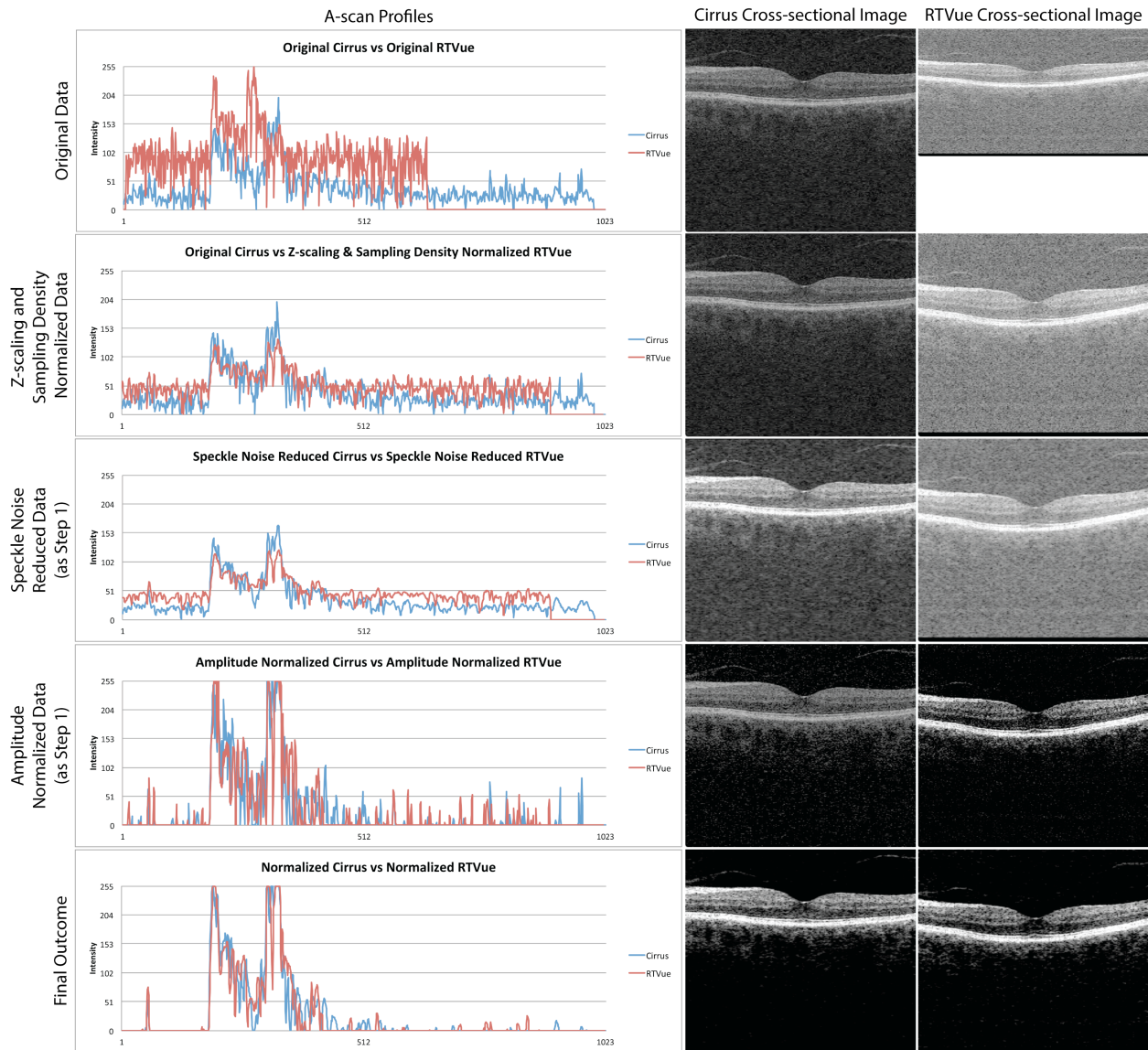
Figure 19. OCT A-scan profiles in the consecutive steps of the signal normalization procedure. First row: Original A-scan profiles from Cirrus (blue) and RTVue (red). For display purposes, the A-scan profile from RTVue was linearly scaled from 12-bit to 8-bit. Second row: Z-scaling and sampling density normalization. Third row: Speckle noise reduction. Fourth row: After amplitude normalization, the distance between noise levels from Cirrus and RTVue equals zero. Last row: The combined signal normalization method. Signal profiles of the devices become similar at the end of the process.

Table 10 summarizes the mean absolute amplitude differences in percentage between two A-scan profiles from two original Cirrus scans, from Cirrus and Z-scaled RTVue scans, from normalized Cirrus and RTVue scans, and from two normalized Cirrus scans. Two A-scan pairs in the temporal quadrant, 2 A-scan pairs in the superior quadrant, 1 A-scan pair in the nasal quadrant, and 2 A-scan pairs in the inferior quadrant were excluded because of misalignment. The mean absolute difference in percentage between Cirrus and RTVue was statistically significantly reduced after signal normalization (12.7 vs 6.2 %, p<0.0001, paired t-test). The mean absolute difference in the amplitude in the percentage between Cirrus and RTVue were also statistically significantly decreased after normalization in all quadrants (p<0.0001, paired t-test).

Table 10. Mean absolute difference in amplitude between A-scan profiles. Absolute difference is in percentage with 95% CI in parentheses.

	Overall (%)	Temporal n = 19	Superior n = 19	Nasal n = 20	Inferior n = 19
Original Cirrus vs Original Cirrus	9.9 (9.6, 10.1)	10.0 (9.4, 10.5)	9.9 (9.3, 10.6)	9.6 (9.2, 10.0)	9.9 (9.4, 10.4)
Original Cirrus vs Z-scaled RTVue	12.7 (12.4, 13.0)	12.8 (12.1, 13.4)	12.4 (11.6, 13.1)	13.0 (12.5, 13.5)	12.6 (12.0, 13.2)
Normalized Cirrus vs Normalized RTVue	6.2 (6.0, 6.4)	6.0 (5.6, 6.4)	6.2 (5.8, 6.7)	6.5 (6.2, 6.9)	6.2 (5.8, 6.5)
Normalized Cirrus vs Normalized Cirrus	6.0 (5.8, 6.2)	5.9 (5.3, 6.5)	6.2 (5.7, 6.7)	5.9 (5.6, 6.2)	6.0 (5.6, 6.3)

After signal normalization, the overall mean absolute difference in percentage between Cirrus and RTVue was statistically significantly smaller compared to the difference between two

Cirrus scans (6.2 vs 9.9 %, p<0.0001, paired t-test), indicating that the signal normalization process successfully reduced the differences, even lower than the level of the intra-device difference. Similar results were found in quadrant analysis: the difference between the Cirrus and RTVue data was statistically significantly smaller compared to the difference between two Cirrus scans in all four quadrants (p<0.0001, paired t-test).

The last row in Table 10 summarizes the residual in percentage between two normalized Cirrus scans. After signal normalization, the residual in percentage between two scans from the same device was also statistically significantly reduced for both mean and all quadrants (p<0.0001 for all comparisons, paired t-test), indicating that the proposed signal normalization method also reduced the difference among the OCT data obtained with the same device. Comparing the residual between normalized Cirrus and RTVue and two normalized Cirrus data, the overall residual was statistically significantly different (6.3 vs 6.0 %, p=0.03, paired t-test). For the quadrant analysis, there was no statistically significant difference in the residual between two normalized comparison pairs except for the nasal quadrant (p=0.0006).

Table 11 and Table 12 present the results divided by the clinical grouping. A statistically significant reduction in the mean absolute difference in percentage between Cirrus and RTVue after signal normalization was observed for both healthy and glaucoma groups for the mean and all quadrants (p<0.0147, paired t-test). Furthermore, after signal normalization, the absolute difference in percentage between Cirrus and RTVue was also statistically significantly smaller as compared to the difference between two Cirrus scans in all comparisons for each group.

Table 11. Mean amplitude absolute difference between A-scan profiles for healthy subjects. Absolute difference is in percentage with 95% CI in parentheses.

	Overall (%)	Temporal n = 13	Superior n = 12	Nasal n = 14	Inferior n = 14
Original Cirrus vs Original Cirrus	9.7 (9.5, 10.0)	9.7 (9.1, 10.2)	9.9 (9.2, 10.6)	9.5 (9.0, 10.0)	9.9 (9.3, 10.5)
Original Cirrus vs Z-scaled RTVue	12.6 (12.3, 13.0)	12.7 (11.8, 13.6)	12.3 (11.4, 13.2)	13.0 (12.3, 13.6)	12.5 (11.8, 13.2)
Normalized Cirrus vs Normalized RTVue	6.3 (6.1, 6.5)	6.1 (5.5, 6.7)	6.3 (5.8, 6.9)	6.5 (6.0, 7.0)	6.1 (5.7, 6.6)
Normalized Cirrus vs Normalized Cirrus	6.0 (5.8, 6.2)	5.9 (5.3, 6.4)	6.3 (5.6, 7.0)	5.9 (5.5, 6.2)	6.0 (5.6, 6.4)

Table 12. Mean amplitude absolute difference between A-scan profiles for glaucoma subjects. Absolute difference is in percentage with 95% CI in parentheses.

	Overall (%)	Temporal n = 6	Superior n = 7	Nasal n = 6	Inferior n = 5
Original Cirrus vs Original Cirrus	10.1 (9.6, 10.6)	10.6 (9.1, 12.1)	10.0 (8.5, 11.5)	9.9 (9.1, 10.6)	9.9 (8.8, 10.9)
Original Cirrus vs Z-scaled RTVue	12.9 (12.3, 13.4)	13.0 (12.1, 14.0)	12.4 (10.9, 14.0)	13.2 (11.9, 14.4)	12.8 (11.6, 14.1)
Normalized Cirrus vs Normalized RTVue	6.1 (5.8, 6.5)	5.8 (4.9, 6.7)	6.0 (5.1, 7.0)	6.5 (5.7, 7.3)	6.2 (5.5, 6.9)
Normalized Cirrus vs Normalized Cirrus	6.0 (5.6, 6.4)	6.0 (4.3, 7.7)	6.0 (5.0, 7.0)	6.0 (5.3, 6.7)	5.9 (5.0, 6.9)

4.1.3   Discussion

In this experiment, we developed a novel signal normalization method to reduce the A-scan profile differences between two SD-OCT devices. The presented method successfully reduced the differences between A-scan profiles from Cirrus and RTVue.

The effect on reducing the residual (in percentage) between Cirrus and RTVue A-scan profiles of speckle noise reduction and amplitude normalization was assessed separately. Each processing step focused on different factors that resulted in the dissimilarity between Cirrus and RTVue signals, and solved them from a different aspect. Speckle noise reduction was applied to eliminate the differences between Cirrus and RTVue A-scan profiles caused by the randomly distributed high spiky signal (considered to be the speckle noise). Amplitude normalization was designed to remove the noise level difference between two devices. Overall, both processing steps significantly reduced the A-scan profile residual with some tested cutoffs. Nevertheless, since each individual processing step focuses on one particular factor, the ability to reduce the residual between Cirrus and RTVue of each one was limited. These two processing steps complement each other in order to achieve the optimized signal normalization.

The optimized cutoff settings for each processing step were different, indicating that the cutoff settings were processing specific. Various cutoffs were used and tested for their ability to reduce the residual between Cirrus and RTVue in speckle noise reduction and amplitude normalization. Different reactions to reduce the residual with various cutoffs were observed between individual processing steps. In order to optimize the signal normalization method, the cutoffs that showed the best results were chosen when combining each processing step to build the final signal normalization method. For speckle noise reduction, the 33rd percentile on the histogram was picked; while for amplitude normalization, the 66th percentile was used as the low

cutoff to remove the noise level difference. There is still some room for further optimization of the method, for example, systematic software training with feedback to find the optimal solution.

We picked the 66th percentile on the histogram as the low cutoff for amplitude normalization as the best setting. Though setting an even higher cutoff for amplitude normalization may have provided smaller residuals, we needed to strike a balance between reducing the profile differences and preserving the actual retinal signals. Based on the statistical analysis of the regular retinal thickness of the entire scan length of the OCT frame, the meaningful retinal signals form approximately one-third of the OCT images. The same analysis results are also applied in traditional and conventional ways of displaying OCT images on the devices using false-color scheme; the lower 66% of signals (or similar cutoff) are usually considered as noise signals and removed, and the entire OCT images are further rescaled so that the details of the retinal tissues can be appreciated in a clearer way so physicians can make clinical diagnoses.

Despite the fact that our results support and work well with the notion of using the 66th percentile as the low cutoff, there is still the possibility that we may discard some actual retinal tissue signals as having intensity weaker than the strong noise signal by cutting off low intensity pixels, which may result in the loss of important information from ocular tissues. In order to dynamically calculate the optimized cutoff for multiple SD-OCT devices that separate true tissue signal from the noise, Huang et al. have developed a method based on histogram density modeling and decomposition.[88] However, they found that the overlap of weak retinal tissue signal and strong noise signal is relatively wide, and results in the limitation that we may lose information from retinal tissues having less reflectivity by cutting off low intensity pixels. Further improvement on this issue is required.

The proposed signal normalization method can be applied to and works for all sampling points in the cube data. However, for validation purposes, the sampling location was chosen to have all the inner retinal layers distinguishable, as the layers merge or disappear in the area close to the foveola, while avoiding the major retinal blood vessels. It is known that blood vessels reflect and block the light signal, which causes shadowing artifacts that obscure the ocular tissue information beyond the blood vessels (Figure 20).[107, 108]
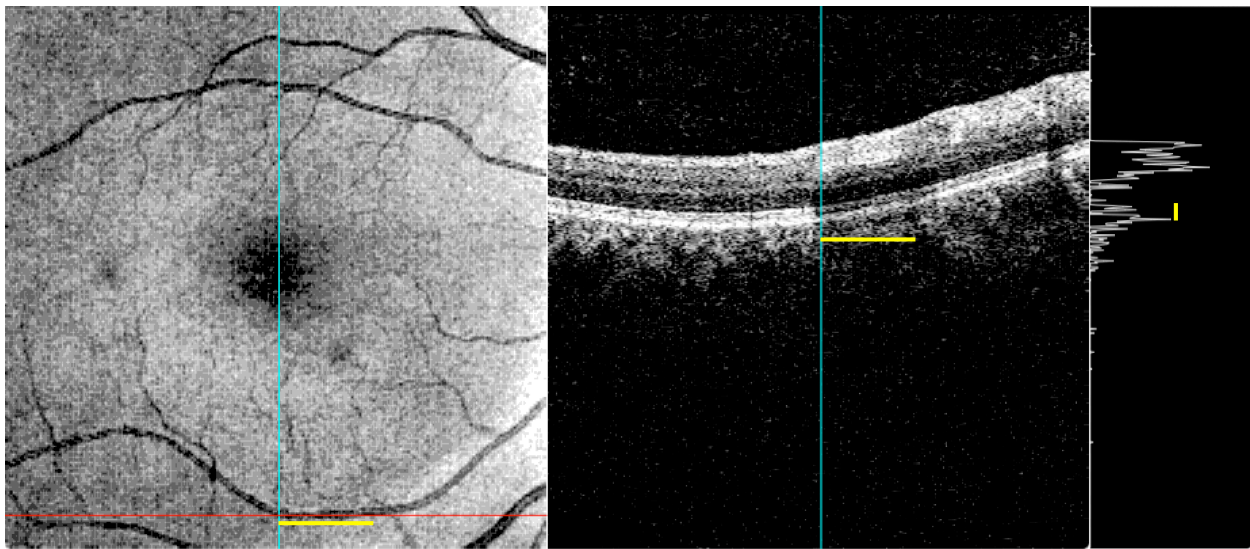
Figure 20. Shadow effect of blood vessel on OCT signal quality. The horizontal red line on the *en face* image (left) indicates the position where the cross-sectional image (middle) was sampled, 2.61 mm away from the foveola. A dramatic drop of the signal quality can be observed at the positions where blood vessel lies (the yellow bar) on the cross-sectional image and the A-scan profile (right), which is sampled from the position of the cyan vertical line.

Other intensity profile normalization methods of OCT images across different eyes using the same devices have been developed to compensate for the RNFL thickness measurement variability caused by inconsistent attenuation of the reflectivity between healthy and diseased eyes, and to increase the sensitivity and specificity of disease detection.[74, 109] Many studies

solved the A-scan profile intensity variation by normalizing the RNFL signal intensity to the brightest layer in the OCT image, usually the RPE, as a preprocessing step before thickness measurement to reduce the variation. Those methods are simple, easy to implement, and can be applied to the entire OCT image. However, they require a robust segmentation algorithm to accurately detect the positions and boundaries of RNFL and RPE, which can be a challenge, especially with coexisting retinal pathology. Another disadvantage of those methods is that they assume that the relative signal intensity of adjacent retinal layers is the same across different eyes and SD-OCT devices. From our observation, A-scan profiles varied substantially among different eyes and SD-OCT devices. Even on the same eye, on different devices, they can significantly differ from each other on intensity proportions among various retinal sub-layers. For example, A-scan profiles from Cirrus and RTVue (the second row of Figure 19) have different contrasts between high and low intensity signal, where the contrast is larger in Cirrus than in RTVue, which results in different responses from the same segmentation algorithm, or different RNFL thickness measurements. In contrast, our method does not require any segmentation prior to the normalization, and has no assumption about the intensity profiles being similar on the same eye across different devices. In addition, our signal normalization works equally well on healthy and glaucomatous eyes, indicating that the proposed method is capable of compensating for signal characteristic differences independent from the pathologic state, where the RNFL signals are generally weaker and show somewhat different A-scan profiles.

Although we only tested the effect of the method with Cirrus and RTVue devices, in principle, this normalization method can be applied to all SD-OCT devices. Further investigation is warranted.

In conclusion, the reported novel signal normalization method successfully reduced the A-scan profile differences between Cirrus and RTVue SD-OCT in healthy and glaucomatous eyes. This signal normalization method would allow the establishment of fundamental signal compatibility among multiple OCT devices, which would make the analysis and measurement results from various devices directly comparable.

4.2   CLINICAL VALIDATION OF THE SIGNAL NORMALIZATION

An engineering validation was conducted to test the signal normalization's ability to reduce the variability of characteristics in OCT signals by measuring the absolute differences in A-scan profile intensity.[103] The results were promising. To move one step further toward our ultimate goal, which is making the outcome measurements directly comparable among multiple OCT devices, a clinical validation was performed to validate the ability of the signal normalization method to reduce the systematic measurement differences among SD-OCT devices. The circumpapillary RNFL thickness measurements obtained from Cirrus and RTVue were compared before and after signal normalization processing.

4.2.1   Methods

Subjects included in this study were recruited at the University of Pittsburgh Medical Center Eye Center Glaucoma clinic (both healthy and glaucomatous eyes), following the same tenets described in Section 2.0. The inclusion criteria were the same as Section 2.1.

The circumpapillary region from all eyes was scanned using Cirrus HD-OCT (software version 5.1) and RTVue (software version 6.1) at the same visit. Scan patterns, which allow the devices to measure the RNFL thickness using their own segmentation algorithms, were used on both devices, namely the Optic Disc Cube 200×200 scan for Cirrus and the RNFL 3.45 Circle scan pattern for RTVue. The details of the protocols were described in Section 2.2.

4.2.1.1 Signal Normalization Processing

The signal normalization was performed as previously described.[103] The normalization process had three stages: 1) z-scaling and sampling density normalization, 2) amplitude normalization, and 3) image quality normalization. Since a modified median filter was applied as part of the preprocessing stage in our segmentation algorithm to reduce the speckle noise,[85] the speckle noise reduction step in the original signal normalization method was removed. On the other hand, as image quality has been known to be an important factor affecting the signal responses to segmentation algorithms and inducing RNFL thickness measurement variability, in order to compensate for the differences in image quality, image quality normalization, which was omitted in the previous experiment, was added after amplitude normalization. For image quality normalization, two methods were applied individually to normalize image quality variation, and thus generated two signal normalization methods: one that used HDR processing at the image quality stage, and one that used HM processing.

Method 1: With HDR Processing

In the first method, HDR processing was applied after amplitude normalization and used to minimize image quality difference. Quality index (QI) was first calculated and used as the image quality index for each OCT image.[87] As the data range of QI was different between

91

Cirrus and RTVue images, QI values were normalized by calculating the percentiles of the QI distribution on Cirrus and RTVue separately. Then the difference in QI percentiles between the matched Cirrus and RTVue images obtained from the same eye was calculated to assess signal quality disparity. Finally, on the histogram of the QI percentile difference, the top and bottom 5 percent of the differences were classified as cases showing substantial QI difference, which became subjects to the image quality normalization.

In each Cirrus and RTVue pair, the image with worse quality was processed with our custom HDR processing (details described in Section 3.3.1) to compensate for poor image quality.[110] In brief, the HDR processing remaps the signal dynamic range in three signal levels (low, medium, and high) separately, and then combines them into one so that OCT retinal signal is enhanced and boosted selectively on poor signal portion of the images.

Method 2: With HM Processing

In the second method, since HM processing has been used to calibrate the differences in intensity contrast and intensity dynamic range between images, in the signal normalization method using HM processing at the image quality normalization stage, amplitude normalization and image quality normalization were combined together and replaced with HM processing so that the final signal normalization method only had two stages: 1) z-scaling and sampling density normalization and 2) HM processing.

Reference Histogram Construction

The HM processing started with a reference histogram construction so that input image histograms could be shaped to a reference histogram. Instead of preparing a specific reference

histogram for each eye, a group reference histogram was constructed. 40 Cirrus OCT images of 40 healthy eyes from the same dataset with a SS larger than or equal to 9 (maximum signal strength is 10) were selected and used to created the reference histogram. The histogram based on the data of the entire 3D cube was generated and the average histogram of the 40 healthy eyes was calculated and used as the group reference histogram for Cirrus data.

For the RTVue data, the reference histogram was constructed using the same Cirrus OCT images from the same healthy subjects. To mimic the circular scan type of RTVue data, the geometric centers of the ONH of the Cirrus images were first automatically determined based on the manually delineated disc margins. The A-scans located on the 3.45 mm circle were sampled to generate the virtual circular Cirrus scans and then the histogram was calculated from the re-sampled circular scans. Finally, the mean of the virtual circular histogram from the 40 Cirrus OCT images was calculated and used as the group reference histogram for RTVue circular scan data.

Histogram Matching

The HM processing was applied to all the Cirrus and RTVue data. Details of the HM processing were described in Section 3.4.2. Instead of using the average intensity from the 3×3 neighbor, the sub-feature we added here was the difference between each sampling point's top and bottom neighbors so that the contrast information could be kept. The idea was based on a Sobel filter,[111] which here we modified into an one-dimensional filter along the axial direction so that the contrast could be kept within each A-scan as shown in Eq. 5-1:

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \qquad \text{(Eq. 5-1).}$$

With the sub-feature, each sampling point has a new value as presented in Eq. 5-2:

$$I_{new} = I_{ori} \times 2 \times DataRange + \left[(I_{bottom} - I_{top}) + DataRange - 1\right] \quad \text{(Eq. 5-2)},$$

where the *DataRange* is 256 for Cirrus data and is 4096 for RTVue data; in order to keep the intensity change direction, instead of using the absolute difference between bottom (I_{bottom}) and top (I_{top}) neighbors, the direction was preserved by shifting the difference with $DataRange - 1$ and thus twice the *DataRange* is multiplied to I_{ori}. The HM processing was performed based on the new intensity and corresponding histogram.

4.2.1.2 RNFL Thickness Measurements

The original machine measured global mean circumpapillary RNFL thicknesses on the original Cirrus and RTVue data were exported from the commercial devices (Comparison I, Table 13). In order to eliminate the measurement differences caused by segmentation algorithm variation and test the hypothesis that applying the same segmentation algorithm can reduce the measurement differences, RNFL thickness was also measured automatically using the same universal RNFL segmentation algorithm of our own design as before (Comparison II).[85] As the word "universal" indicates, our segmentation software is able to open, read, and perform retinal layer segmentation on various SD-OCT data with the same core segmentation algorithm, unlike algorithms integrated in the commercial devices which have an optimized approach and parameters targeting the signal characteristics of a specific SD-OCT device, and thus may not generate equally good segmentation results when processing OCT data from a different SD-OCT device. In Comparison III, RNFL thickness was measured using the same universal algorithm, but with parameters tuned specifically to Cirrus and RTVue images in order to assess the effect of fine-tuning the universal algorithm. Finally, RNFL thickness was measured after signal

normalization (for both HDR and HM processing), using the universal segmentation algorithm without any specific tuning (for Comparison IV, the parameter settings were the same as in Comparison II). The segmentation performance was subjectively evaluated for any potential erroneous border detection. Image data were excluded if the images demonstrated one or both of the following: (1) apparently inaccurate border detection for more than a consecutive 15% or an additive 20% of the total image or (2) the borders of the RNFL collapsed, meaning that the RNFL thickness was recorded as a string of zeros for at least 10 consecutive points.

Table 13. 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	Algorithm parameters only optimized for Cirrus and applied this algorithm to both original Cirrus and RTVue data
Comparison III	Original signal	Algorithm parameters optimized for both Cirrus and RTVue separately and applied to both original Cirrus and RTVue data
Comparison IV	Normalized signal	Algorithm parameters optimized for Cirrus and applied this algorithm to both normalized Cirrus and RTVue data

4.2.1.3 Statistical Analysis

In order to appropriately handle the comparison between RNFL thickness measurements from Cirrus and RTVue with multiple measurements of the same RNFL thickness from data including both eyes from the same subject, we constructed a comprehensive measurement error model. This measurement error model describes how the true unknown RNFL thickness of each eye is

linked to the measurements from each device and the processing method, and provides calibration equations to delineate the relationship between Cirrus and RTVue for different comparison methods.

The simplified basic measurement error model is given by Eq. 5-3:

$$x_{ij} = \alpha_i + \beta_i \mu_j + \varepsilon_i \qquad \text{(Eq. 5-3)},$$

where μ_j indicates the unknown true RNFL thickness for the jth eye, x_{ij} indicates an RNFL thickness observation measured by device i (Cirrus or RTVue) for eye j, α_i and β_i describe the bias (systematic error) introduced by device i, and ε_i denotes a random error whose distribution describes the imprecision for each device. Based on the measurement error model, the calibration equation between two devices for each comparison method can be derived as Eq. 5-4:

$$E[x_R] = \left[\alpha_R - \frac{\beta_R}{\beta_C} \alpha_C \right] + \frac{\beta_R}{\beta_C} E[x_C] \qquad \text{(Eq. 5-4)},$$

where E denotes the expectation operator (which averages out the random error) and C stands for Cirrus while R stand for RTVue. When the ratio of two device slopes (β's) equals one, the calibration line is parallel to the no-bias line, $E[x_C] = E[x_R]$, and the bias is considered to be a constant bias and equal to the horizontal or vertical distance between the calibration line and the no-bias line.

SEMs (Figure 21) were used to estimate the parameters in the measurement error model and further derive parameters for the calibration equations. The R environment and language for statistics (version 2.13.1)[112] with OpenMx (version 1.1.2-1818)[113] and merror (version 1.0)[114] were used to describe the SEMs. Full information on the maximum likelihood was used to estimate the measurement error model parameters.
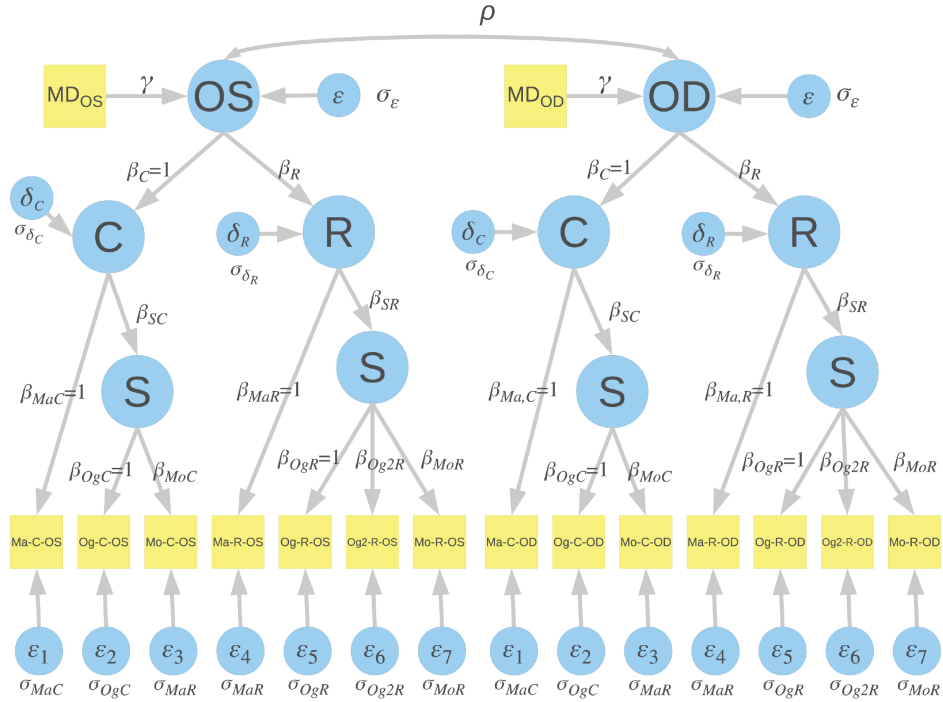
Figure 21. Path diagram of the SEMs used for statistical analysis.

Furthermore, to assess the effect of our signal normalization method in reducing the systematic difference between Cirrus and RTVue over a wide range of disease severity (measured using the visual field mean deviation (MD) value), linear mixed effect models were constructed to estimate the relationship between the differences in RNFL thickness between two devices to the MD value.

4.2.2 Results

One hundred and nine eyes from 59 subjects were included in this study. Subject demographics and clinical characteristics are presented in Table 14. Disease severity, as measured by the visual field MD, ranged from -9.23 to 2.13 dB, including healthy subjects as well as early and moderate glaucoma subjects.

Table 14. Subject demographics. Data are expressed as the mean ± SD and 95% CI in the parentheses.

Subject Demographics	
Male / Female	16 : 43
Age (years)	61.67 ± 8.0 (59.60, 63.75)
Visual field mean deviation (MD) (dB)	-0.76 ± 1.90 (-1.13, -0.40)

Table 15 shows the global mean circumpapillary RNFL thicknesses from Cirrus and RTVue data measured using five different methods, the systematic differences in RNFL thickness measurements between Cirrus and RTVue, and the slope and intercept values of the corresponding calibration lines. The RNFL thicknesses between Cirrus and RTVue were statistically significantly different before normalization regardless of the choice of segmentation algorithms (device built-in or our custom design, Comparison I to Comparison III). Before signal normalization, there were significant differences in RNFL thickness measurements between Cirrus and RTVue both from the original device outputs (Comparison I; mean absolute difference 10.6 μm, p<0.05, Figure 22A) as well as when using the same segmentation software regardless of using the same parameters for Cirrus and RTVue or optimizing the parameters for Cirrus and RTVue separately (Comparison II and III; 18.1 μm and 10.9 μm, both p<0.05, Figure 22B and C, respectively). After signal normalization with HDR processing, although the RNFL thickness showed a non-constant difference between devices (Comparison IV-HDR; Figure 22D), the difference was reduced substantially. The difference between Cirrus and RTVue was statistically significantly reduced by signal normalization for the eyes with RNFL thicker than 62.4 μm according to Cirrus device measurements (dotted blue vertical line in Figure 23), representing 95% of the studied population. The mean absolute difference between Cirrus and RTVue for eyes with a Cirrus RNFL thickness larger than 62.4 μm was 2.95 μm, which was

calculated by averaging the absolute difference between the two devices in that range. On the other hand, after signal normalization with HM processing, the measurement differences between Cirrus and RTVue were significantly reduced across the entire thickness range (mean absolute difference 0.5 μm, p>0.05, Figure 22E). The largest difference between Cirrus and RTVue after signal normalization with HM processing was 2.5 μm, which appeared when the RNFL thickness was 48.7 μm according to Cirrus device measurements, which was also the minimum RNFL thickness in our dataset.

Table 15. Summary of RNFL thickness differences using four comparison methods. Global mean circumpapillary RNFL thickness measurements and systematic measurement differences between Cirrus and RTVue, along with the slope and intercept values of the corresponding calibration line, are presented. Data are expressed as the mean ± SD and 95% CI are shown in the parentheses. *: Non-constant difference ranging from -0.1 to 5.0 μm.

	Cirrus (μm)	RTVue (μm)	Diff (μm)	Slope	Intercept
Comparison I	82.7 ± 12.0 (80.4, 84.9)	93.0 ± 12.7 (90.6, 95.5)	10.6 (9.8, 11.4)	1.1 (1.0 1.1)	6.2 (-1.8, 13.3)
Comparison II	96.8 ± 15.1 (93.9, 99.6)	114.5 ± 17.2 (111.3, 117.8)	18.1 (16.3, 20.0)	1.1 (0.9, 1.2)	13.5 (-3.3, 28.5)
Comparison III	96.8 ± 15.1 (93.9, 99.6)	107.4 ± 16.1 (104.3, 110.4)	10.9 (9.2, 12.6)	1.0 (0.9, 1.1)	13.5 (-2.0, 27.3)
Comparison IV-HDR	97.7 ± 15.0 (94.9, 100.6)	99.6 ± 14.8 (96.8, 102.4)	*	0.9 (0.8, 1.1)	9.2 (-6.3, 22.7)
Comparison IV-HM	96.6 ± 14.3 (93.8, 99.4)	96.7 ± 13.9 (94.0, 99.4)	0.5 (-1.3, 2.3)	1.0 (0.8, 1.1)	5.7 (-12.4, 21.0)

(A)

(B)
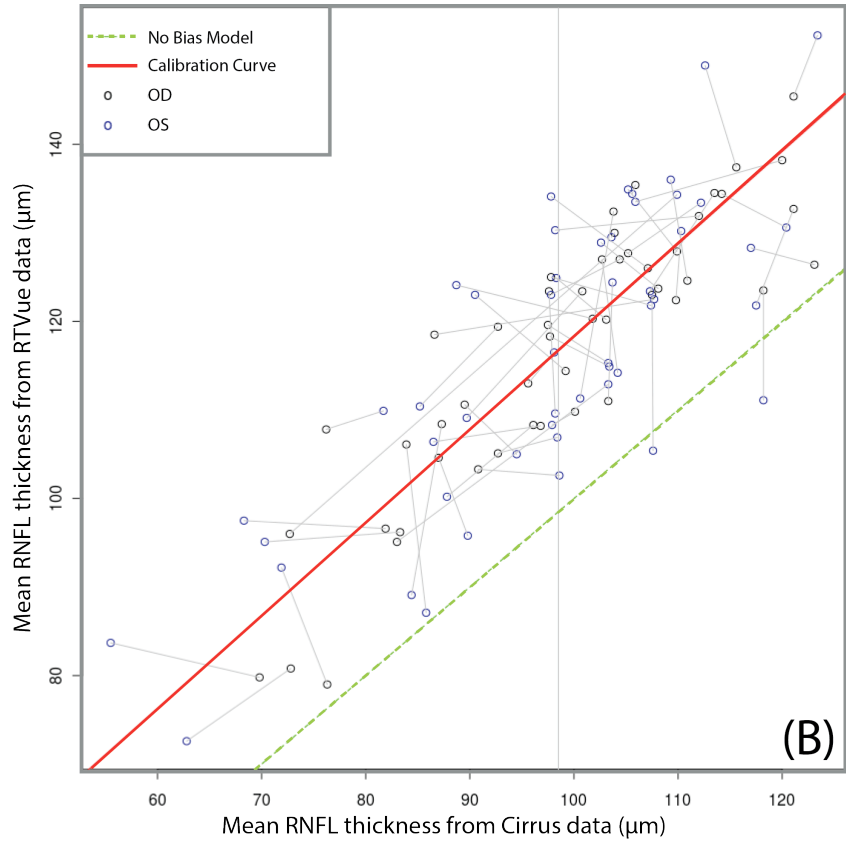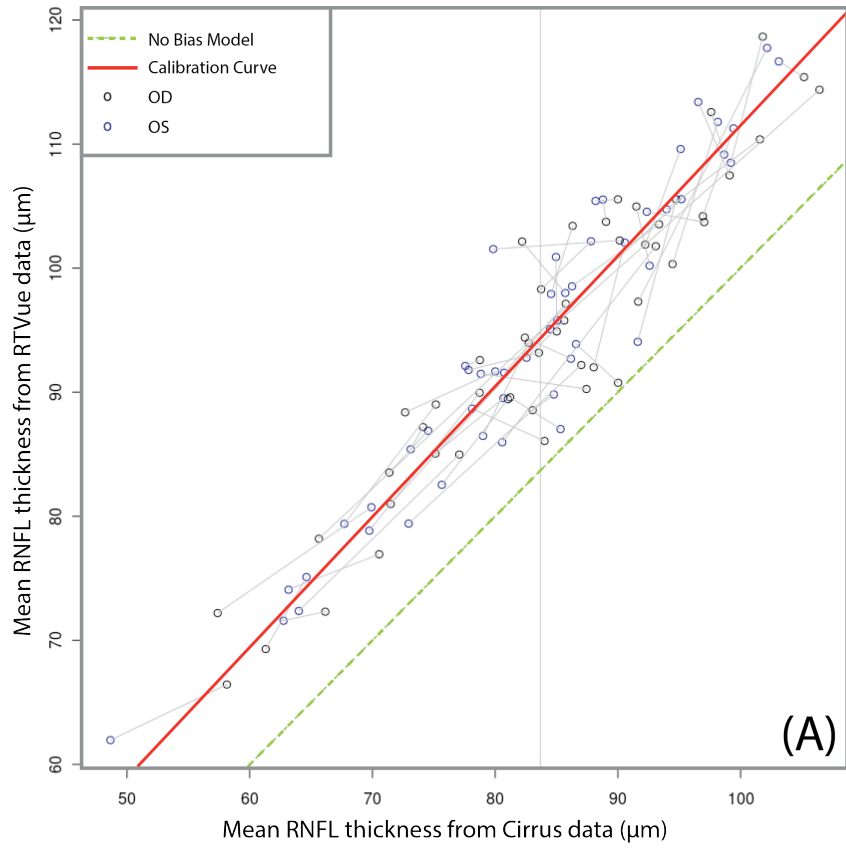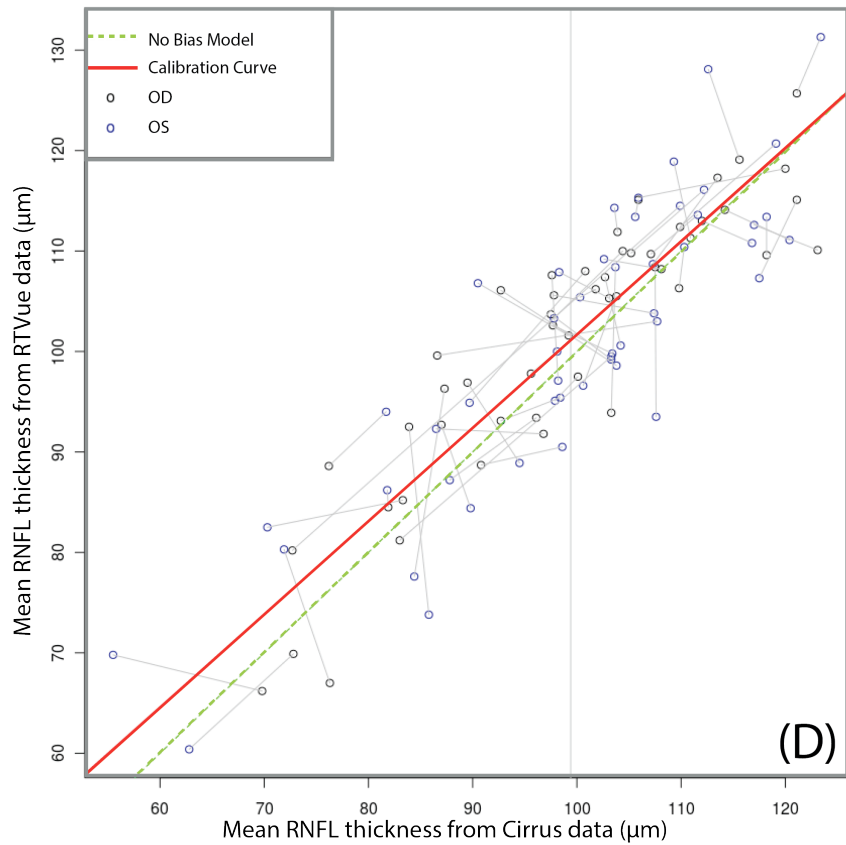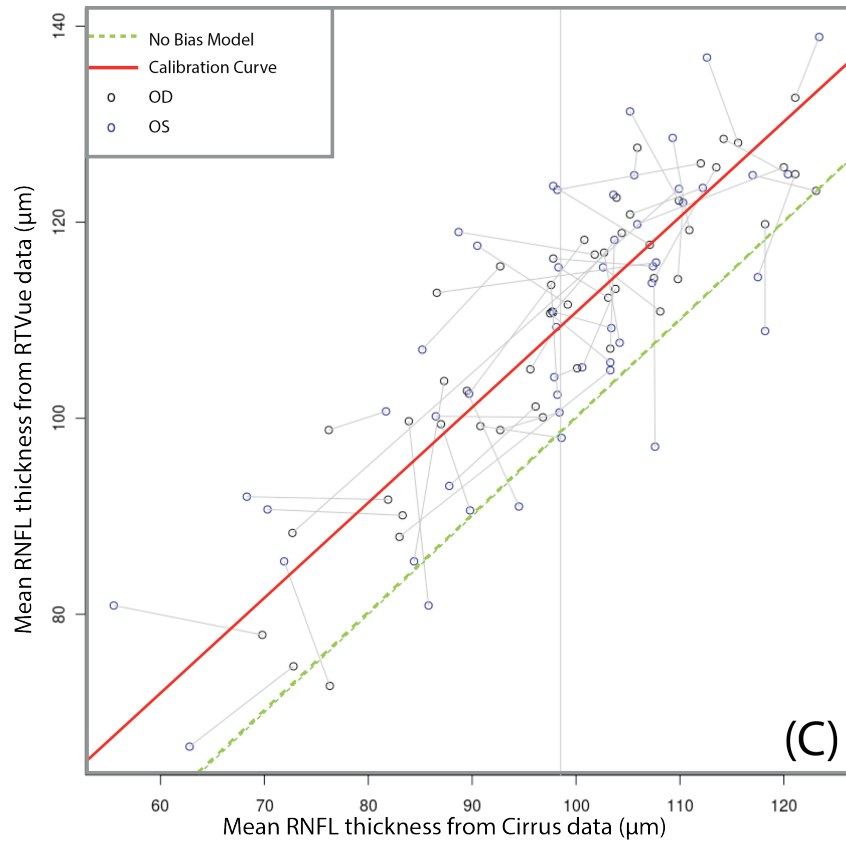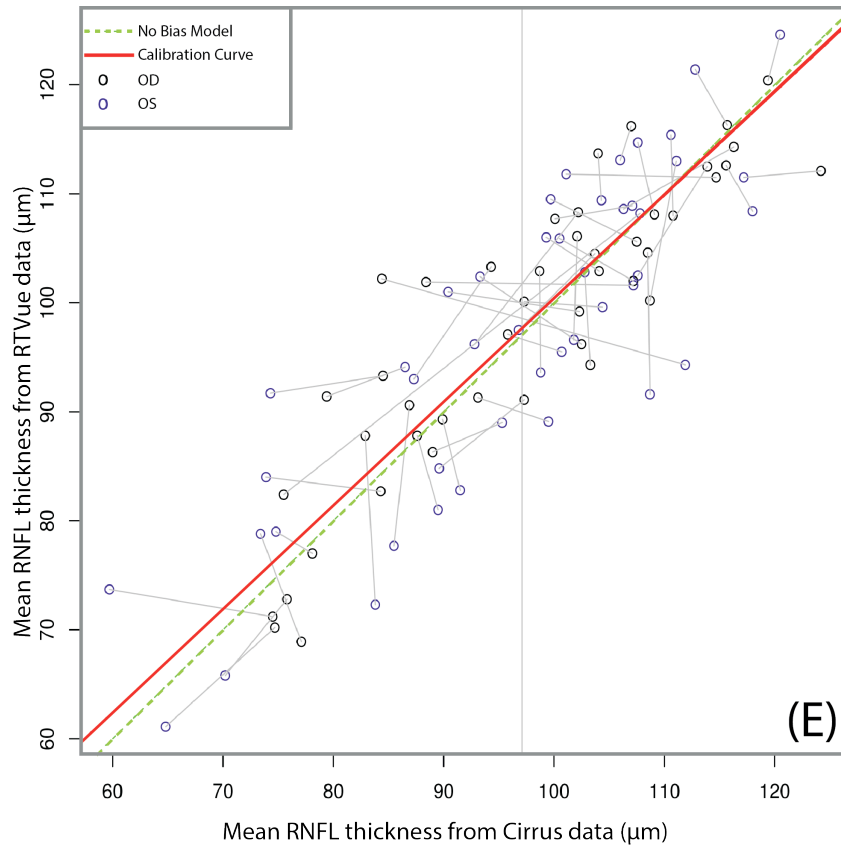
Figure 22. Scatter plots of RNFL thickness measurements with five comparison methods. (A)-(C) Comparison I-III, (D)-(E) comparison IV with HDR processing and HM processing, respectively. The calibration curve (red line) and no-bias curve (green line) were drawn on each plot. Vertical gray line indicates the average of the RNFL thickness measured from Cirrus data, and the constant differences between two SD-OCT devices were measured as the distance between red line and green line at this point. Paired eyes from the same subject are connected by gray lines. See **Table 13** for definition of the comparisons.
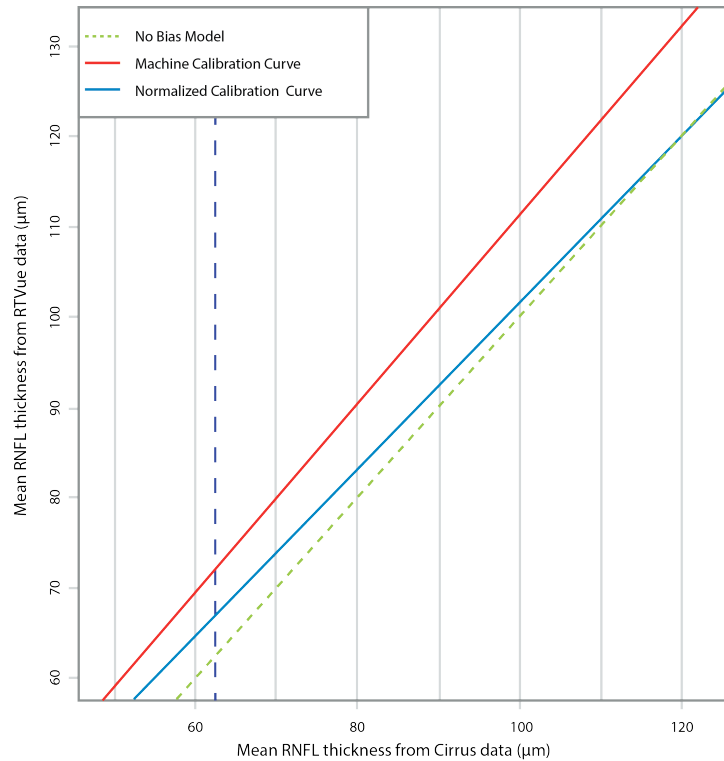
Figure 23. Calibration curves between Cirrus and RTVue RNFL thickness measurements. Blue line: after signal normalization with HDR processing, red line: from original machine outputs, and green dotted line: no-bias curve. Note how the calibration curve after normalization (blue line) is closer to the no-bias curve (green dotted line), showing less bias overall. However, the blue line is not parallel to the green dotted line, indicating that the systematic measurement difference depends on the measured thickness. The dotted blue vertical line is at the threshold of RNFL thickness where differences between devices below this level are statistically significant.

The relationship between the RNFL thickness differences (RTVue − Cirrus) and visual field MD was also analyzed. Since the effect of the present signal normalization on eyes with a Cirrus-measured RNFL thinner than 62.4 μm were not significant, six eyes with such condition were excluded for this analysis. The RNFL difference showed no significant correlation with the visual field MD both before and after normalization (correlation coefficient 0.24 vs -0.34 μm/dB, respectively, p>0.30), indicating that the residual difference between two devices was independent from disease severity before and after normalization.

103

4.2.3　Discussion

The systematic measurement differences between two commercial SD-OCT devices, Cirrus and RTVue, were statistically significantly reduced to the level of measurement variability within devices after processing with the developed signal normalization. Along with the results presented in Section 4.1, it is not only indicated that the proposed signal normalization is able to reduce the residuals between A-scan profiles, but it is also capable of minimizing the measurement differences between OCT devices, which further improves the comparability of OCT data measurements among machines.

Two methods were included in the signal normalization process, HDR and HM processing. When processed using the signal normalization with HDR processing, it successfully reduced the differences for most of the cases (95%) where the Cirrus RNFL thickness was larger than 62.4 μm. This encompasses a wide range of subjects including early and moderate glaucoma participants, along with healthy subjects. However, the results also demonstrate that the present method had limited ability to reduce the systematic measurement differences for cases where the Cirrus RNFL thickness is less than or equal to 62.4 μm, where advanced glaucoma subjects stand. The behavior of the systematic measurement difference can be observed from Figure 23. The absolute difference in RNFL measurements between Cirrus and RTVue after normalization was always smaller than the difference between the original device outputs. Although the calibration curve after normalization (blue line) in Figure 23 is closer to the no-bias curve (green line), showing less difference overall, it is not parallel to the no-bias curve, indicating that the systematic measurement difference depends on the measured thickness. We assumed that the lack of ability to reduce the measurement differences when Cirrus RNFL

thickness was less than or equal to 62.4 μm comes from an insufficient number of observations of advanced glaucoma subjects, and further investigation with more observations of severe glaucoma cases is warranted.

When processed using the signal normalization with HM processing, the systematic measurement differences were significantly reduced across the entire thickness range (including all clinical groups: healthy, early, moderate, and advanced glaucoma subjects). The largest difference between Cirrus and RTVue was 2.5 μm after normalization, less than the largest difference when using HDR processing (5.0 μm). This suggests that signal normalization with HM processing performed better than normalization with HDR processing. One advantage of using HM processing was the generality. When performing HDR processing, image pairs with significant image quality differences were carefully selected and then the one with lower image quality was subjected to HDR processing, so that we could avoid signal saturation in high image quality scans. This selection scheme enhanced the poor quality images without boosting high quality images, however, and also induced a selection bias (which may cause a proportional bias in measurements between devices). On the contrary, HM processing can be applied to all image data without any selection, compensating for the quality differences between image pairs, and thus it successfully eliminated the selection bias and improved the measurement differences between SD-OCT devices across the entire thickness range. Another advantage gained from HM processing was the efficiency. Three-stage signal normalization was integrated into two-stage normalization, which makes the proposed method simpler and more efficient. One potential limitation of HM processing is its requirement of a reference histogram. As described in Section

3.4, the reference histogram can be generated from a group of images instead of individual reference image, however, it is still a speculation that group reference works with all the images including both healthy and diseased eyes. Further investigation is warranted.

Previous studies showed that RTVue-measured RNFL thicknesses were thicker than the corresponding Cirrus measurements.[58, 59] The results of our analysis of the devices' original outputs agreed with the previous findings. Heussen et al. suggested in a recent study that similar RNFL thickness measurements could be generated by both manually segmenting and correcting the outer retinal boundary to a standardized reference location.[115] Their results support our first hypothesis that while scanning the same eye, OCT signals from different devices contain the same information though the signal characteristics vary because of different device settings and thus react differently to the same segmentation algorithm. However, only correcting the boundary position still cannot resolve all the systematic difference in RNFL measurements, and it is not practical to manually correct segmentation in regular clinical settings.[115] The same results were also observed in our study. Even using the same segmentation software, RNFL thickness measured on the RTVue images still presented thicker measurements than when measured on the Cirrus images, which further implies that the factors causing this systematic difference in RNFL thickness is not only due to the use of a different segmentation algorithm but also due to various signal characteristics.

The systematic RNFL measurement difference between Cirrus and RTVue was 10.6 μm from the original devices outputs, and 18.1 μm when processing both the original Cirrus and RTVue image data with our universal segmentation software without optimizing the parameters for each SD-OCT device separately. The increased systematic measurement difference indicated that simply processing OCT data from different SD-OCT machines with the same segmentation

algorithm cannot reduce the systematic measurement differences between SD-OCT devices, but makes the differences even larger. The results also proved our hypothesis that OCT data from different SD-OCT devices have different signal characteristics so that they react differently to the same segmentation algorithm.

After fine-tuning the parameters in our universal segmentation software for Cirrus and RTVue separately, the systematic measurement difference became 10.9 μm. With the optimization, the systematic measurement difference decreased to the same level as the one obtained from machine outputs, 10.6 μm. This result was expected since tuning parameters in the same algorithm for each specific SD-OCT device worked similarly to using different algorithms, which were optimized for specific SD-OCT devices and would present the best performance for the image captured from the specific device. However, software optimization did not fix the systematic measurement difference between Cirrus and RTVue data. Therefore, a different approach other than adjusting the segmentation algorithm is needed to solve this problem.

With the present signal normalization method, RNFL thickness from the two devices could be reduced to the inherent device measurement variability level and become directly comparable. By unifying the sampling density in the axial direction using z-scaling and sampling density normalization, normalization of the pixel dynamic range, and compensating for image quality differences, the proposed normalization method succeeded in transforming OCT signals obtained with one device into virtually similar signals obtained with the other device. Although the systematic differences in RNFL measurement between Cirrus and RTVue could not be reduced to a statistically significant level with RNFL thickness thinner than 62.4 μm, the largest difference between two devices after normalization was 5.0 μm, which was within the inherent device measurement variability.

It was interesting that the RNFL thickness measurement differences between devices were independent from the disease severity. With thinner RNFL on glaucomatous eyes, one may expect a smaller difference if the effect is proportional. But instead, the results suggest that the effect is more of a fixed bias regardless of the disease status. It is likely this bias is stemmed from the characteristic difference in the slope of the intensity profiles at the inner and outer borders of the RNFL. We hypothesize that normalizing such intensity profile characteristics would further reduce the systematic difference in OCT measurements. Further investigation is needed.

Although we only tested the effect of the signal normalization method on reducing the systematic RNFL thickness measurement differences with Cirrus and RTVue devices, in principle, this normalization method can be applied to all SD-OCT devices. Further investigation is warranted.

In conclusion, our signal normalization method successfully reduced the systematic difference in RNFL thickness measurements between Cirrus and RTVue to the level of the device's reported inherent measurement variability. This enables the direct comparison of RNFL thicknesses obtained from multiple devices, and would broaden the use of OCT technology in both clinical and research applications.

4.3   SIGNAL NORMALIZATION BETWEEN SINGLE-FRAME AND AVERAGED-FRAME OPTICAL COHERENCE TOMOGRAPHY IMAGES

We have demonstrated that the proposed signal normalization method is capable of unifying the variation of signal characteristics, reducing the residuals between two A-scan profiles, and

minimizing the systematic measurement differences between SD-OCT devices. So far, all the validations focused on the comparability between single-frame OCT images, like Cirrus and RTVue. In this section, we extend the ability of our signal normalization method to normalize between single-frame and averaged-frame images.

The Spectralis system has a hardware eye-tracker built in to resolve eye movement artifacts or blinking during scanning. Two beams of light reach the target eye simultaneously; one beam for tracking and the other beam for OCT scanning, so that Spectralis can repeatedly scan at the same location and apply the OCT signal averaging on the fly. With eye tracking and signal averaging, Spectralis produces much clearer, better contrasted, and more detailed images of the retinal layers. However, it also induced signal characteristic differences as compared to signal-frame OCT images. The previously developed signal normalization method successfully reduced the measurement differences between SD-OCT devices providing single-frame images, but was ineffective in normalizing signals between single-frame and averaged-frame images. The challenge for us here is how to compensate for the signal characteristics variation between single-frame and averaged-frame image. To solve this problem, we developed a novel virtual averaging method and applied it to single-frame image data only in order to minimize the differences between single-frame and averaged-frame OCT data, which are Cirrus and Spectralis.

4.3.1 Methods

Twenty-one healthy subjects volunteered to participate in this prospective cross-sectional study. The right eye from each subject was used in the study. The inclusion criteria were described in Section 2.1.

For each participant, both macular and ONH regions were scanned using two commercially available SD-OCT devices during the same visit, where Cirrus device (software version 6.1) was used to acquire single-frame image data while Spectralis device (software version 1.5) had an eye-tracking system and provided averaged-frame image data. The details of the scan protocols were described in Section 2.2.

4.3.1.1 Virtual Averaging Algorithm Development

To mimic the acquisition of frame-averaged OCT data, we first simulated the deviation in the x and y direction while Spectralis scans at the same location. For each sampling voxel, one neighboring voxel was randomly selected from the 3×3 neighborhood voxels located on the same z-position (including the center voxel) following a 2D Gaussian random distribution, where the closer the voxel was to the center, the higher its possibility of being selected as a candidate (as Figure 24, Step 1 shows). Then we simulated the variation of the voxel value while imaging the same position multiple times by adding a random Gaussian deviation to the selected voxel value (Figure 24, Step 2). This process (deviation in x and y direction and in voxel value, namely Step 1 and 2 in Figure 24) was repeated 15 times for each voxel, and the average of the outcomes were used to replace the original value (Figure 24, Step 3).
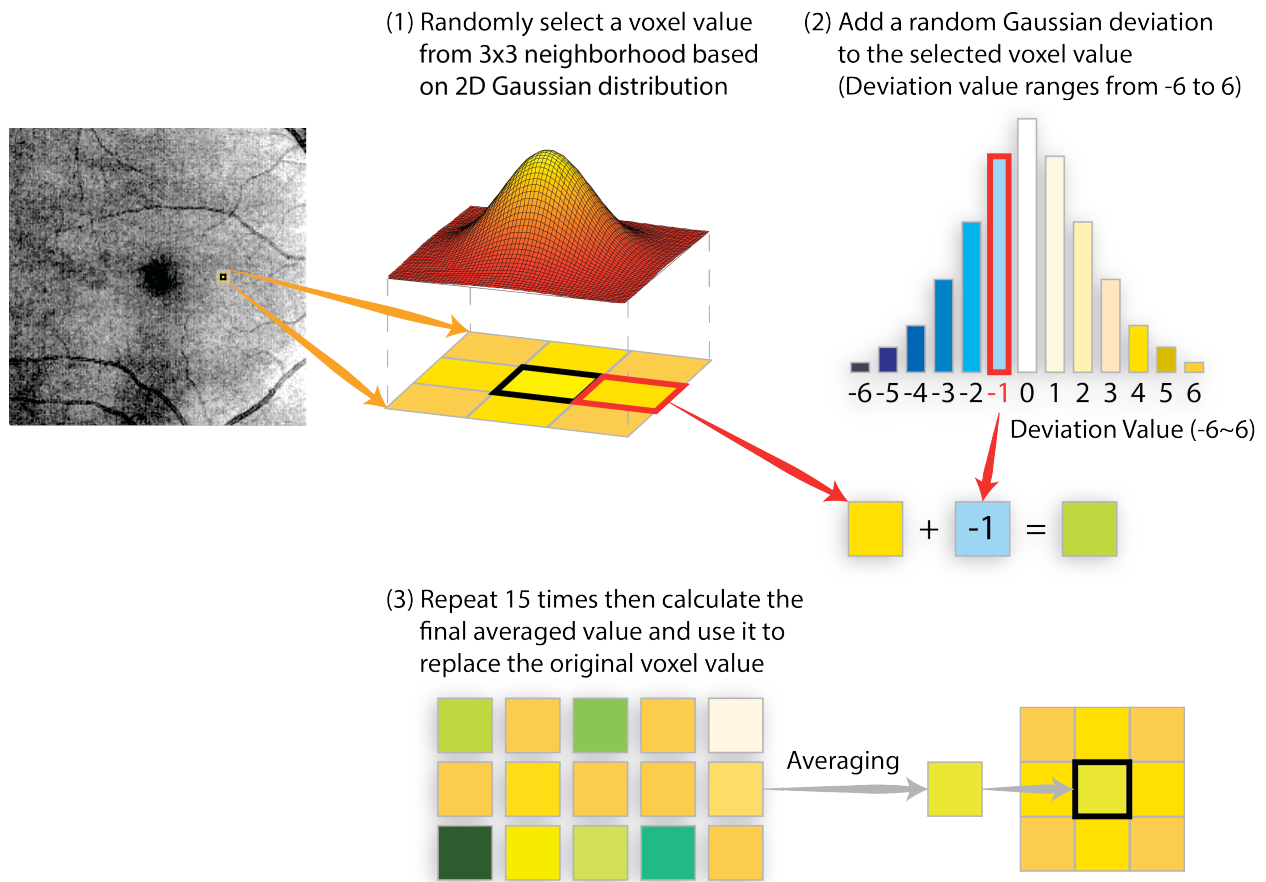
Figure 24. Processing flow of virtual averaging.

4.3.1.2 Virtual Averaging Effects Assessment

For the validation of virtual averaging, we first assessed its effects on OCT image quality enhancement. To test the image enhancement ability, single-frame OCT data (Cirrus data) were processed with the virtual averaging method and the outcomes were evaluated and compared with averaged-frame OCT data (the corresponding Spectralis data) subjectively and objectively.

Subject and Objective Assessment

For subjective assessment, a subjective image quality evaluation based on the image appearance in terms of tissue contrast, the smoothness of the tissues, and the visibility of intra-retinal layers were assessed by two observers (CLC and HI). For objective assessment, SNR and CNR (as described in Section 3.2.2) were calculated to evaluate the image enhancement effect.

In addition to conventional image quality metrics, the distance between the end of visible nasal RNFL and the foveola (dNFL) was measured to assess the effect on improved retinal layer visibility quantitatively. The end of visible nasal RNFL (as indicated by the orange arrow in Figure 25) is usually judged in doubt because the end of the nasal RNFL is too thin to see as the RNFL merges into GCL at the fovea. The definition of dNFL is presented in Figure 25 as the horizontal distance between the blue and orange arrows, where the blue arrow points to the foveola position while orange arrow points to the end of visible nasal RNFL.
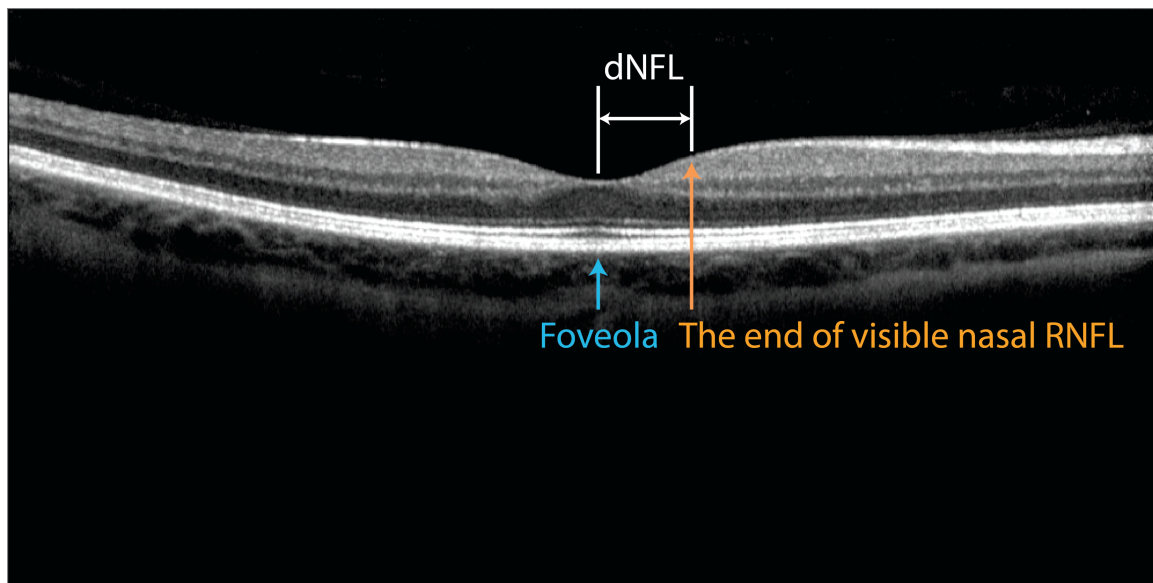
Figure 25. Distance between the end of visible nasal RNFL and the foveola measurement. Blue arrow points to the foveola position while orange arrow points out the end of visible nasal RNFL. dNFL is calculated as the horizontal distance between the blue and orange arrows.

Statistical Analysis

Paired t-tests were used to analyze the image quality metrics (SNR and CNR) improvement between the original and virtually averaged images as well as the differences in dNFL between single-frame and averaged-frame data before and after processing.

4.3.1.3 Virtual Averaging Application in Signal Normalization

Virtual averaging was then employed in the signal normalization and tested for its compensating effects on the signal characteristic differences as well as reducing tissue thickness measurement differences between single-frame and averaged-frame images.

Signal Normalization

The signal normalization was modified based on the previously reported method.[103, 116] The normalization process had two disparate stages for Cirrus and Spectralis data, as presented in Figure 26. Spectralis data (averaged-frame data) were first processed with Z-scaling and sampling density normalization, while Cirrus data (single-frame data) were processed with virtual averaging in the first step. Since 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 removed. After the first step, both Cirrus and Spectralis data were processed with amplitude normalization to match the intensity dynamic range.
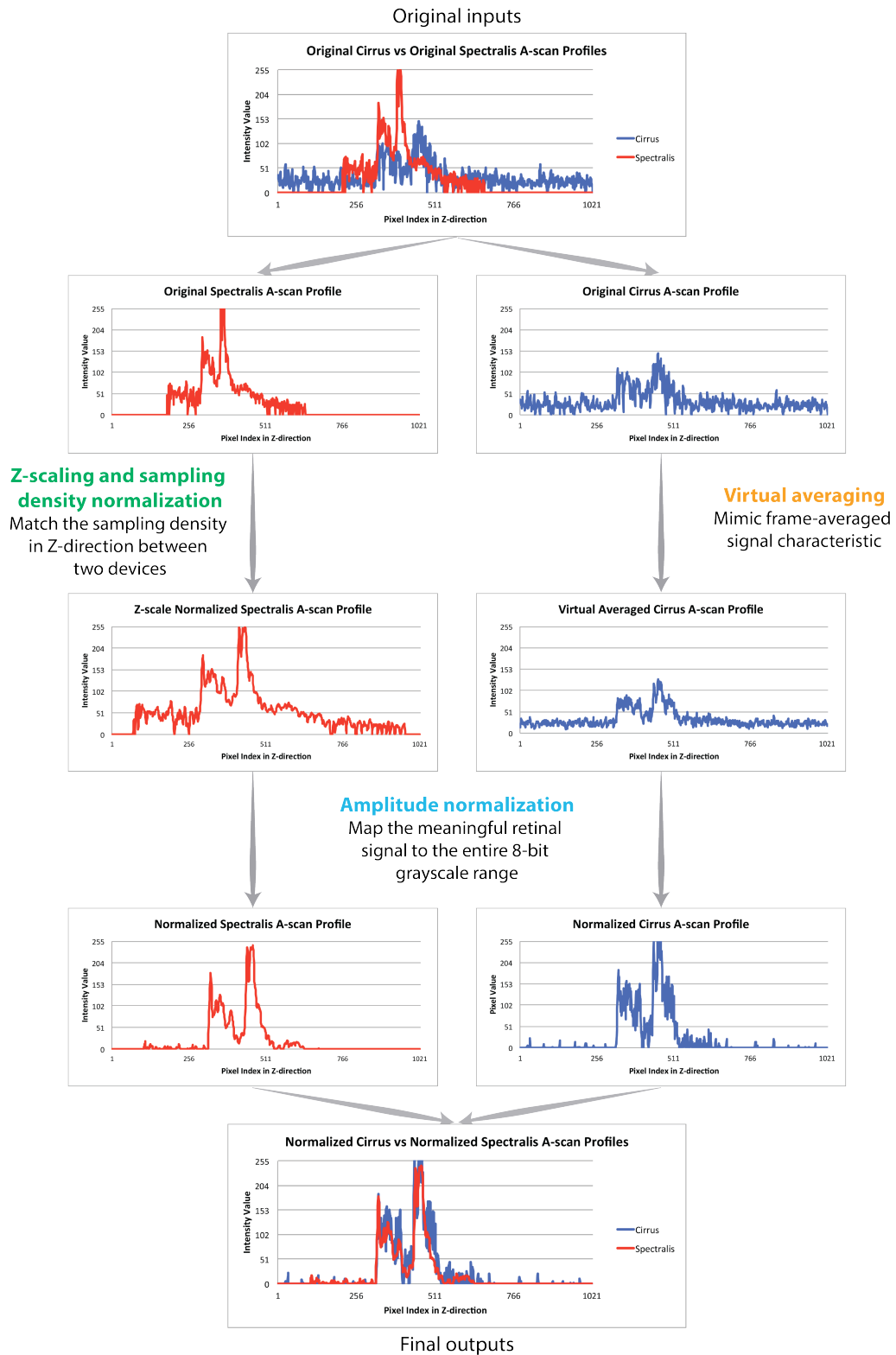
Figure 26. Flow chart of signal normalization for single-frame and averaged-frame data.

OCT Thickness Measurement Comparison

The same thickness measurement comparison methods were used as described in Section 4.2.1.2 and Table 13. Both the total retinal thicknesses of nine macular sections and the global mean circumpapillary RNFL thicknesses were collected. In order to measure in the same region for sectoral macular total retinal thicknesses, the foveola position was manually selected by looking for the largest separation between the junction of IS/OS of the photoreceptors and the RPE as appearing on the horizontal and vertical cross-sectional B-scans (as described in Section 4.1.1.1, shown in Figure 16). An Early Treatment Diabetic Retinopathy Study (ETDRS) pattern was then applied to measure the total retinal thickness in nine sectors.

Statistical Analysis

As mentioned in Section 4.2.1.3, SEMs were used to analyze the absolute difference in the circumpapillary RNFL thicknesses and the sectoral total retinal thicknesses in the macular region between Cirrus and Spectralis from the original machine outputs, from our software outputs before and after signal normalization.

4.3.2   Results

Twenty-one right eyes from 21 healthy subjects were recruited in this study. They included 9 males and 12 females. The mean age of the healthy group was 34.3 ± 11.5 years old. The averaged visual field MD was -0.6 ± 1.1 dB.

Subject and Objective Assessment

Subjectively, all processed images showed notable improvement in image quality and bore clear resemblance to active tracking averaged Spectralis images after virtual averaging. Figure 27 presents an example of the cross-sectional images in the macular region from the original single-frame Cirrus image (left), processed Cirrus image (middle), and averaged-frame Spectralis image (right). As the example shows, the external limiting membrane was hard to differentiate from the signal pattern in the outer nuclear layer (ONL) in the original Cirrus image, but became clearly visible and easy to trace after processing (red arrowhead). Moreover, the contrast between retinal layers became more apparent and the continuous inner border of the IPL (yellow arrowhead) became easily distinguishable after virtual averaging.
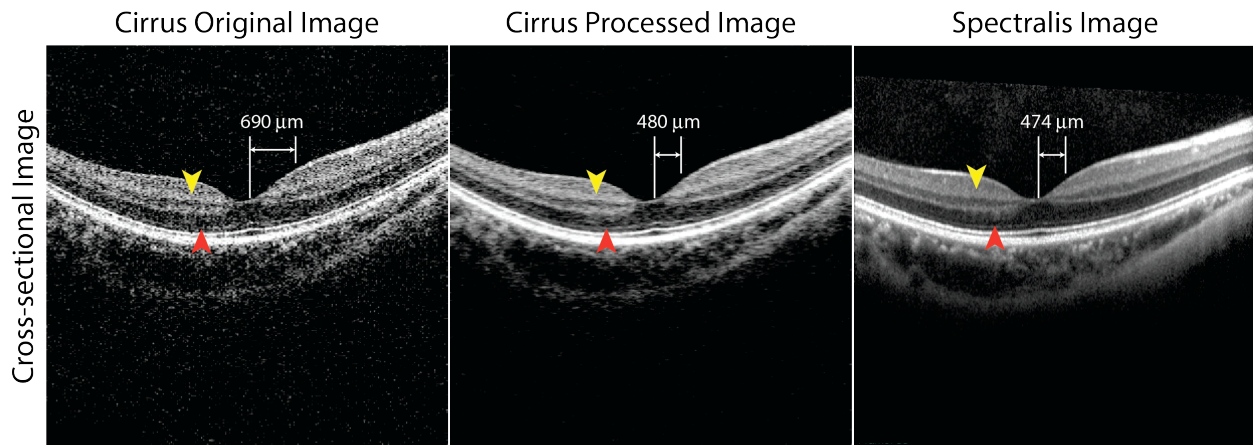
Figure 27. An example of virtual averaging. Processed Cirrus data show notable improvement in signal quality and retina; cross-sectional image looks similar to averaged-frame Spectralis image. In the processed Cirrus image, ELM (red arrowhead) and the continuous inner border of IPL (yellow arrowhead) become clearly visible.

In the objective assessment using image quality metrics, the mean SNR and CNR were significantly improved after virtual averaging (SNR: 30.5 vs 47.6 dB, CNR: 4.4 vs 6.4 dB, original vs processed, p<0.0001, paired t-test). As for quantitative analysis on the dNFL parameter, dNFL were significantly different between Cirrus and Spectralis before processing (681.4 vs 446.5 μm, original vs Spectralis, p<0.0001, paired t-test), but after virtual averaging there was no significant difference in dNFL between Cirrus and Spectralis (442.9 vs 446.5 μm, processed vs Spectralis, p=0.76, paired t-test).

Signal Normalization and Compensating Tissue Thickness Measurement Differences

For sectoral macular total retinal thicknesses, significant systematic differences were detected in all sectors between Cirrus and Spectralis on both device outputs (Comparison I, Table 16, p<0.0001) and between our universal software measurements before normalization (Comparison II and III, Table 16). 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 16). 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).

Table 16. Results of total macular retinal thickness analyses between Cirrus and Spectralis. Sectoral macular total retinal thickness measurements and systematic measurement differences between Cirrus and Spectralis, using four comparison methods were summarized. 95% CI are shown in the parentheses. *: Statistically significant differences were detected between Cirrus and Spectralis.

		Comparison I (Device outputs)			
		Cirrus (µm)	Spectralis (µm)	Mean Absolute Differences (µm)	P-value
Outer	Temporal	258.3 (253.0, 263.6)	281.7 (276.4, 286.9)	23.3 (20.6, 26.1)	
	Superior	278.0 (272.3, 283.6)	298.9 (292.7, 305.1)	21.0 (18.7, 23.2)	
	Nasal	297.8 (290.2, 305.4)	315.9 (308.1, 323.7)	18.1 (15.3, 21.0)	
	Inferior	266.6 (260.8, 272.4)	287.4 (281.0, 293.8)	20.8 (18.3, 23.3)	
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)	
	Nasal	324.8 (318.9, 330.6)	346.9 (341.5, 352.3)	22.1 (19.1, 25.2)	
	Inferior	317.9 (311.6, 324.1)	341.6 (336.2, 346.9)	23.7 (21.0, 26.5)	
	Center	259.0 (252.7, 265.2)	275.0 (269.0, 280.9)	16.0 (13.5, 18.5)	

Table 16 (continued).

		Comparison II (Our custom algorithm with same parameters)			
		Cirrus (µm)	Spectralis (µm)	Mean Absolute Differences (µm)	P-value
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 (28.5, 285.0)	31.7 (12.5, 50.8)	0.34
		Comparison III (Our custom algorithm with fine tuned parameters)			
		Cirrus (µm)	Spectralis (µm)	Mean Absolute Differences (µm)	P-value
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*

Table 16 (continued).

		Comparison IV (After signal normalization)			
		Cirrus (µm)	Spectralis (µm)	Mean Absolute Differences (µm)	P-value
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
	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

For the circumpapillary RNFL thickness, statistically significant differences were found between Cirrus and Spectralis in the original device outputs and when using our custom segmentation software (Comparison I to III, Table 17). Although the differences found in the device outputs were significant, the detected absolute differences (3.6 µm) were within the expected device measurement variability and were clinically non-significant (Comparison I).[64, 117, 118] Applying the same segmentation algorithm (our custom universal segmentation software) did not help reduce the systematic differences in RNFL thicknesses, but made the differences larger (Comparison II and III, Table 17). After signal normalization, no systematic measurement differences were detected between Cirrus and Spectralis (Comparison IV, Table 17). 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 vs 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 artifacts.

Table 17. Results of RNFL thickness measurements between Cirrus and Spectralis. Global mean circumpapillary RNFL thickness measurements and systematic measurement differences between Cirrus and Spectralis, using four comparison methods were summarized. 95% CI are shown in the parentheses. *: Statistically significant differences were detected between Cirrus and Spectralis.

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

4.3.3 Discussion

In this experiment, a virtual averaging method was developed and employed in our signal normalization to improve the comparability of non-averaged non-tracking OCT images and active tracking averaged frame OCT images. By resampling voxels within a 3×3 neighborhood, adding a Gaussian deviation multiple times, and then calculating the average, the proposed

method successfully mimicked the way active tracking averaged-frame devices acquire images, which reduced the measurement differences between single-frame and averaged-frame OCT data and further improved the image quality of single-frame non-tracking OCT data.

We assumed the deviation of the incident light beam was caused by the relocation of the camera as well as the variation in signal intensity from the same location caused by the dryness of the cornea, speckle noise, and other factors following Gaussian distributions. That means the closer the voxel is to the center voxel, or the more moderate the change to the voxel value, the higher the possibility the voxel or the deviation has of being selected. In this way, the output of each re-sampling process would not select a voxel too far away from the center voxel or change the voxel value dramatically, therefore tissue structural information from the original OCT image can be kept after the averaging process.

By adding Gaussian deviation and repeating the process (Step 1 and 2 in Figure 24) multiple times, the outcome of virtual averaging showed much less background noise and speckle noise in the retinal signal and thus strongly improved the image quality and intra-retinal layer contrasts. Retinal tissues like ELM and the end of nasal side RNFL, which cannot be identified clearly in the single-frame images, became clearly visible and easier to delineate after virtual averaging. By enhancing the visualization of single-frame OCT images, the proposed method may help detect the fine structural changes in those originally obscure tomographic features and improve the interpretation and assessment of the progression of pathologies; furthermore, it may enable detailed retinal structure studies on images that previously fell short because of image quality, and may also enable more robust and finer retinal tissue segmentation.

Besides being able to enhance image quality, the virtual averaging can be used to normalize the OCT signal between single-frame images and averaged-frame 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, it would be expected that the systematic measurement differences in macular total retinal thickness between these two devices could be minimized by applying a universal segmentation algorithm and adjusting the foveola position. However, our results showed that even using the same segmentation algorithm, there were still significant differences in the macular total retinal thickness between Cirrus and Spectralis (Comparison II and III, Table 16), and it was indicated that the signal normalization process was still required. After applying signal normalization and using the universal segmentation algorithm, the systematic measurement differences between single-frame and averaged-frame OCT data were successfully reduced and the clinical measurements of macular total retinal thickness from them were made directly comparable.

The circumpapillary RNFL thickness showed a significant difference in the original device outputs between Cirrus and Spectralis. Though the difference did not reach clinical significance as reported in the literature,[64, 117, 118] the detected statistically significant differences in circumpapillay RNFL from the device output indicate a consistent trend of Spectralis circumpapillary RNFL being thicker than the Cirrus measurements. Similar to the results found between Cirrus and RTVue, the systematic differences in circumpapillary RNFL thickness were not reduced by applying the same segmentation algorithm (our universal segmentation algorithm), but became even larger, indicating that signal normalization is needed to minimize the measurement differences. Despite circumpapillary RNFL thickness not needing any post hoc processing to make it clinically comparable between Cirrus and Spectralis, the circumpapillary RNFL thickness measurements trend between Cirrus and Spectralis was not

found 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 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 increasing scanning time, bridges the gap between single-frame and averaged-frame images, and furthermore, makes both qualitative and quantitative assessments between single-frame and averaged-frame OCT images directly comparable.

5.0   QUALITATIVE ASSESSMENT OF SIGNAL NORMALIZATION

The primary application of OCT in ophthalmology is the qualitative evaluation of disease status from cross-sectional images. However, the variability in pixel intensity, ocular tissues contrast, sampling density in the axial direction, image quality, and OCT image noise level may influence the interpretation of OCT images and may induce a serious clinical challenge when clinicians want to compare images from different OCT devices. In previous quantitative assessments, the proposed signal normalization method has presented its ability to minimize the systematic measurement differences and enable direct measurement comparisons among OCT devices, both between single-frame OCT devices and between single-frame and averaged-frame OCT machines. Now, we investigate its ability to reduce the discrepancies in OCT image appearance among OCT devices. Although subtle changes exist in the OCT images scanned using the same OCT device, we assumed larger differences would be present in the OCT images taken with different OCT machines. We hypothesized that by applying our signal normalization technique the similarity in image appearance among OCT devices could be attained. A qualitative validation was conducted to assess the effect of the proposed method on reducing the variation of image appearance among SD-OCT devices by subjectively evaluating of the image similarity before and after signal normalization.

125

5.1 METHODS

This was an observational cross-sectional study. Subjects recruited in this study were collected at the University of Pittsburgh Medical Center Eye Center, including healthy eyes as well as eyes that have previously been shown to have glaucoma, AMD (both dry and wet AMD), diabetic retinopathy (including proliferative diabetic retinopathy (PDR), non-proliferative diabetic retinopathy (NPDR), and diabetic macular edema (DME)), macular hole, and cystoid macular edema (CME). Multiple retinopathology was included to test if the proposed signal normalization method works regardless of pathology. The tenets and inclusion criteria were described in Section 2.0 and Section 2.1.

Both macular and ONH regions from all eyes were scanned using at least two out of three SD-OCT devices (listed as following) at the same visit: Cirrus (software version 6.1), RTVue (software version 6.1), and Spectralis (software version 1.5). Macular and ONH scans were acquired. The details of the scan protocols were described in Section 2.2.

5.1.1 Signal Normalization Processing

Signal normalization methods developed in Chapter 3 and Chapter 4 were applied to all the OCT images. Z-scaling and sampling density normalization and image quality normalization were applied to RTVue and Spectralis data to convert the data into Cirrus equivalent data format. For the original Cirrus and normalized RTVue data, virtual averaging and amplitude normalization were applied to mimic averaged-frame OCT data.

5.1.2 Subjective Evaluation

For subjective evaluation, a questionnaire was created and used to validate the effect of our signal normalization on reducing the variation in image appearance among OCT devices. The exemption of informed consent from survey participants was approved from the University of Pittsburgh Review Board and ethics committee. This study followed the tenets of the Declaration of Helsinki and was conducted in compliance with the Health Insurance Portability and Accountability Act.

The questionnaire contained 30 sets of images (30 questions). Each set contained two cross-sectional images extracted from the same location acquired from the same eye scanned at the same visit with any combination of Cirrus, RTVue, and Spectralis (Cirrus vs RTVue, Cirrus vs Spectralis, RTVue vs Spectralis, and two images from the same device). Observers were asked to evaluate the similarity of the image appearance for the two displayed images (as shown in Figure 28) based on 1) the contrast between the retinal signal and the background noise, 2) the contrast between adjacent retinal layers, and 3) the textures or patterns of RNFL and RPE (the two brightest retinal layers) according to their visual experience. The similarity was recorded in a Likert-type five point scale ranging from one point to five points, where one meant least similarity between the two images and five meant excellent similarity as listed below: 1) Significantly different (0-20% similarity), 2) Somewhat different (20-40% similarity), 3) Cannot decide if it is similar (40-60% similarity), 4) Looks similar (60-80% similarity), and 5) Nearly identical (as if taken by the same device) (80-100% similarity). When judging the similarity, the differences in retinal axial location in the scanning window, retinal orientation, and pathological contexts should not be taken into account.
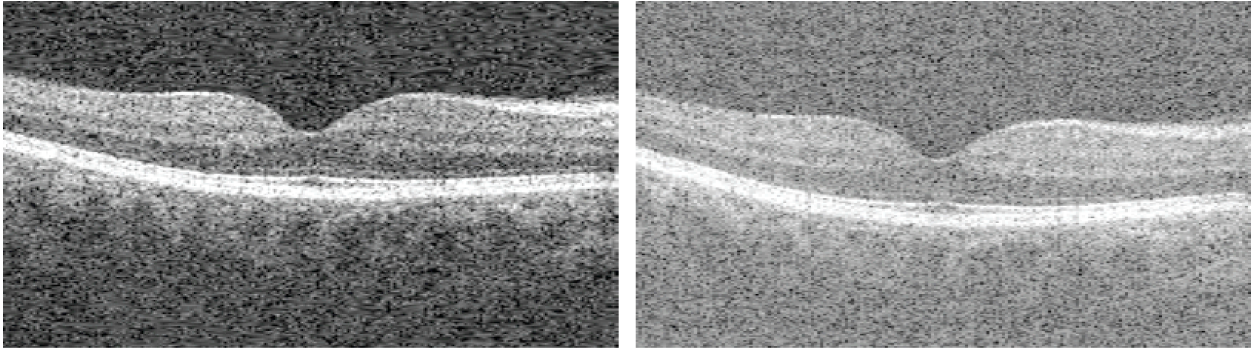
Figure 28. An example of the "significantly different" image pair.

The 30 questions consisted of 24 testing questions and 6 control questions. For the control questions, two displayed images were from the same device (Cirrus vs Cirrus, RTVue vs RTVue, and Spectralis vs Spectralis), either before or after signal normalization. The scores of the 6 control questions were used to establish the scores of similarity within the same device and used as references. The 24 testing questions were from 12 eyes with different SD-OCT devices (4 Cirrus vs RTVue, 4 Cirrus vs Spectralis, and 4 RTVue vs Spectralis) both before and after signal normalization, so that the effect of reducing the discrepancies in image appearance could be assessed. Image pairs with different pathologies before and after processing were displayed in a random and masked manner.

Eye care specialists of various levels including residents, community ophthalmologists, attendings in glaucoma and retina specialties in ophthalmology, optometrists, medical students, medical imaging device technicians, and researchers having experience with OCT were invited to participate in the study. An electronic invitation was sent out first, and if the recipient agreed to participate, a face-to-face meeting was scheduled to complete the questionnaire. No identifiable information of any sort was collected.

128

An additional five sets of images were presented as examples before the actual survey started. Proper decisions were demonstrated to the participants in order to establish a common similar judging standard. The questionnaire is attached in Appendix A.

5.1.3 Statistical Analysis

Ordinal mixture effects models were constructed to investigate the effects on reducing the differences in image appearance and improving the image similarity among OCT devices. The changes in the histograms of the testing questions were analyzed by the ordinal mixture effects models before and after signal normalization. Cumulative link mixed model was used to estimate the parameters for the ordinal mixed effects models in order to assess the effects.

5.2   RESULTS

Thirty-one eyes from 31 healthy subjects, 15 eyes from 15 glaucoma subjects, and 25 eyes from 18 retinal pathology subjects were recruited in this study. The mean age for each group was 40.9 ± 15.4, 65.9 ± 9.1, and 70.8 ± 9.0 years old. The averaged visual field MD was -0.47 ± 1.27, -3.14 ± 0.12, and -8.06 ± 9.7 dB, respectively. Among them, 3 healthy, 2 glaucomatous, 4 AMD, 4 macular hole, 2 diabetic retinopathy, 1 CME with NPDR, and 2 CME eyes were used to create the subjective evaluation questionnaire.

A total of 29 observers participated in the survey. They were 6 ophthalmologists, 5 faculties in Department of Ophthalmology at University of Pittsburgh School of Medicine, 4 residents in ophthalmology, 1 medical student, 5 OCT technicians, and 8 researchers whose research fields are related to OCT.

Figure 29 presents the descriptive summaries for the responses before and after signal normalization. The responses are summarized in histograms. The vertical axis shows the histogram in percentage while the horizontal axis shows the similarity scale, with 1 indicating the least similar and 5 indicating the most similar. Comparing the histograms before and after signal normalization, the similarity responses shifted from dissimilar to similar, suggesting that after signal normalization, the similarity between OCT images was improved.
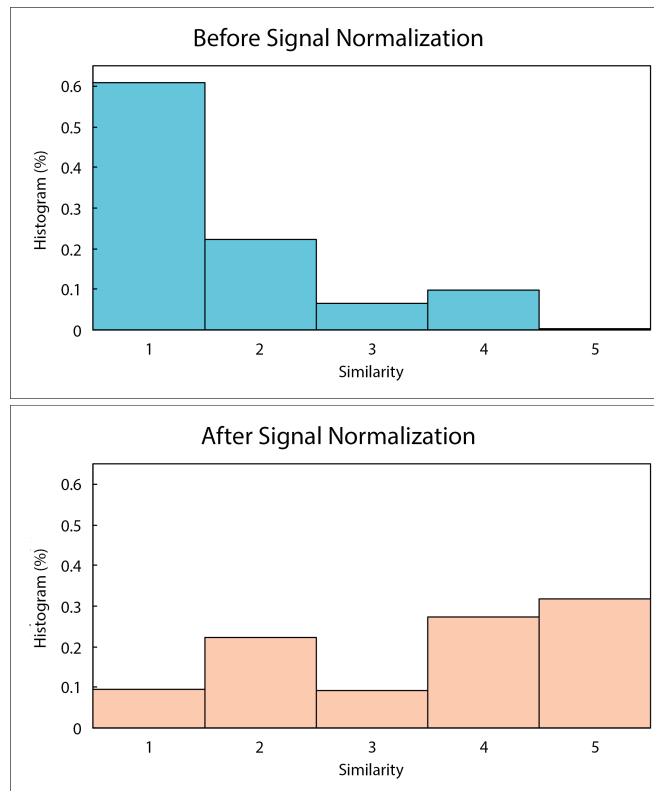

Figure 29. Overall similarity comparison before and after signal normalization. Subjective evaluation results: overall similarity distribution (presented as histogram in percentage) before and after signal normalization.

Figure 30 also shows the descriptive summaries, broken down to the individual comparison groups: Cirrus vs RTVue, Cirrus vs Spectralis, and RTVue vs Spectralis. The same trend was observed in the individual groups: the responses shifted in a positive direction (from dissimilar to similar), further indicating that the signal normalization was able to increase the similarity between OCT images regardless of comparison groups. Another observation was that the image similarity increased after signal normalization to varying degrees among groups, which can also be observed from Figure 30, where group Cirrus vs RTVue showed the highest improvement in similarity, followed by group Cirrus vs Spectralis, and then group RTVue vs Spectralis.
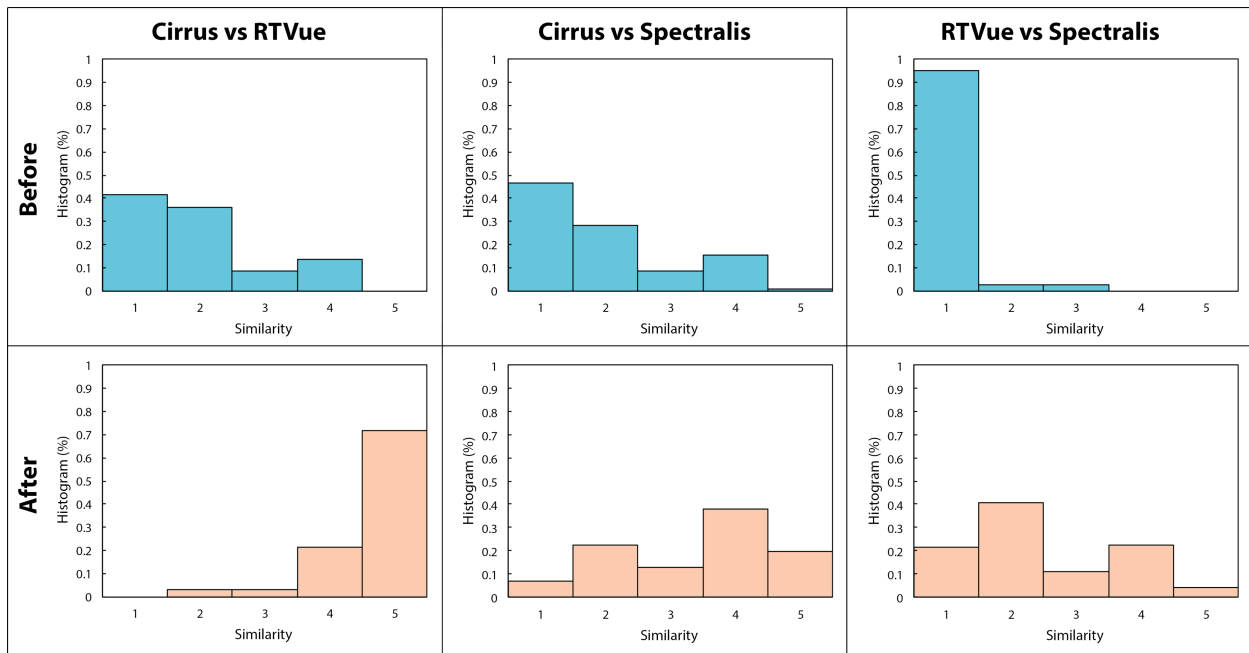

Figure 30. Group similarity comparison before and after signal normalization. Subjective evaluation results: similarity distribution (presented as histogram in percentage) for each comparison group.

131

Table 18 summarizes the estimated parameters for the ordinal mixed effects model. Odds ratios (image similarity of images after signal normalization divided by the similarity before signal normalization) were used to analyze the effect of the signal normalization on improving the similarity between OCT images. An interaction between groups was detected, and thus an interaction model was added in the ordinal mixed effects model. Statistically significant improvement in image similarity was detected both overall and for the individual comparison groups after signal normalization. Varying degrees of the effects depending on the comparison combination were observed, with Cirrus vs RTVue showing the strongest effect (odds ratio in log scale = 6.1, p<0.0001), followed by Cirrus vs Spectralis (2.9, p<0.0001), and then by RTVue vs Spectralis (1.2, p=0.009), the same as our observation from descriptive summaries (Figure 30).

Table 18. Statistical analysis results of the subjective evaluation. *: Statistically significantly different compared to the combination of Cirrus with RTVue before signal normalization.

		Odds Ratio (log scale)	**P-value**
Before	**Cirrus vs RTVue (Reference)**	---	---
	Cirrus vs Spectralis	-0.1	0.82
	RTVue vs Spectralis	-4.0	<0.0001 *
After	**Cirrus vs RTVue**	6.1	<0.0001 *
	Cirrus vs Spectralis	2.9	<0.0001 *
	RTVue vs Spectralis	1.2	0.009 *

5.3 DISCUSSION

A subjective evaluation was conducted through a survey to validate the effects of our signal normalization on reducing the dissimilarity in image appearance among OCT devices. The results showed that the proposed signal normalization method statistically significantly improved the image similarity, in terms of contrasts between retinal layers, textures in ocular tissues, and overall image appearance regardless of pathology.

The successful reduction in the contrast between adjacent retinal layers can be attributed to the normalization in amplitude and intensity dynamic range. By minimizing the differences in the noise level and optimizing the dynamic range to the meaningful retinal signal, the noise signals were suppressed and the actual retinal signals were rescaled and mapped to the same range, which improved the contrasts between retinal signal and background noise, and between adjacent retinal layers, and thus substantially reduced the discrepancies.

Another important factor was the virtual averaging. By mimicking the image acquisition of averaged-frame images, the virtual averaging bridges the differences between single-frame and averaged-frame OCT images. The process not only decreased the quantitative measurement differences, but also improved the qualitative comparison between single-frame and averaged-frame OCT image data. Before processing, the textures or patterns in the retinal layers and ocular tissues, taking RNFL and RPE for example, looked more granular in single-frame image. After virtual averaging, the retinal layers became smoother, clearer, and showed more detailed structural information. Therefore, the signal normalization with virtual averaging was able to reduce the discrepancies in the texture and patterns of retinal layers and enable direct comparisons between single-frame and averaged-frame OCT image data.

Image similarity clearly increased after signal normalization for each individual group, but with different degrees of effects. The combination of Cirrus and RTVue showed the strongest effect, followed by the combination of Cirrus and Spectralis, and finally the combination of RTVue and Spectralis. It was interesting that the signal normalization method successfully made the image appearance comparable between Cirrus and RTVue, and substantially reduced the discrepancies in image appearance between Cirrus and Spectralis, but when it came to convert normalized RTVue (in the equivalent Cirrus data format) to a frame-averaged OCT data, the effects were not the same, though a statistically significant improvement was still observed. The combination of RTVue and Spectralis showed the least similarity before signal normalization while the combination of Cirrus and RTVue and Cirrus and Spectralis were at the same level of similarity (absolute odds ratio difference: 0.1, p=0.82). The differences in similarity among comparison groups may indicate the limited ability of our signal normalization method, or there are other factors resulting in OCT signal characteristics variability. Further investigation is warranted.

A better way to conduct the survey may be testing the effect of the individual aspects (the contrast between retinal signal and background noise, the contrast between adjacent retinal layers, and the textures in the retinal layer) separately so that we can better understand in what aspect the proposed signal normalization has more power. However, it would not be realistic to have such a time consuming survey, especially for busy clinicians. It was a compromise we chose to make in order to maximize the number of clinicians able to participate in the study.

In conclusion, the qualitative validation showed that the proposed signal normalization method is not only able to minimize the systematic differences, but also able to improve the image similarity among SD-OCT devices. By improving the similarity in image appearance

among SD-OCT devices, signal normalization allows direct comparison of OCT images among various instrument, which would broaden the use of OCT technology in both clinical and research applications.

6.0   DISCUSSION

In this dissertation, three objectives were achieved to minimize the systematic discrepancies in OCT signals and the differences in OCT data measurements among three SD-OCT devices. A novel signal normalization method was proposed and developed to unify OCT signal characteristics discrepancy in terms of axial scaling, sampling density, intensity dynamic range, and histogram statistics. For quantitative assessment, the residual between OCT A-scan profiles from different devices (between Cirrus and RTVue) were significantly reduced through bilinear up-sampling, context-aware speckle noise reduction, histogram-based tone mapping, and signal strength normalization (HDR and histogram matching). In the outcome results, the residual among different devices was reduced at least to the level of the observed differences within the same device. The systematic measurement differences in RNFL and retinal thickness measurements across OCT devices are significantly reduced to the level of expected measurement variability in both the peripheral and macular regions (including nine ETDRS sectors as well as the global mean total retinal thickness). For qualitative assessment, the similarity of cross-sectional images among various OCT devices is significantly improved in terms of the contrast between retinal signal and background noise, the contrast between adjacent intra-retinal layers, the texture or pattern in the retinal signal, and the overall image appearance.

Although in this study we only tested the effect of the proposed method with Cirrus, RTVue, and Spectralis devices, in principle, the signal normalization can be applied to all SD-OCT devices since no specific optical characteristics from devices or manufacturers have been used as a prior knowledge for the signal normalization method.

One of the main clinical applications of the signal normalization is in longitudinal studies using multiple OCT devices along the long observation period. In glaucoma practice, long-term quantitative follow-up (trend and change analysis) is usually hard to establish because of the inevitable utilization of various devices. With the proposed method, OCT image data that were previously excluded because of the usage of different OCT devices (in the situation that patients move from one clinic to another clinic) or the change/upgrade of device models and software can be included to set up a longer observation. Furthermore, OCT data acquired using different devices can be used, compared, and analyzed together in the multicenter study. In this way, diverse, large, and long-term follow-up cohorts can be established, which would improve the understanding of the properties of glaucoma, increase the sensitivity and specificity of disease detection, and yield a better method of disease management.

In addition to the effect of reduced OCT signal discrepancies and minimized measurement differences and image inconsistences, the idea of a uniform OCT data format was thought of with the developed signal normalization method. A uniform OCT data format would allow the usage of a universal OCT image browser for both visualization and analysis purposes. No specific image browser will be needed for physicians and clinicians to use OCT image browsing/analysis. No particular settings will be required for OCT images coming from different OCT devices. Physicians and clinicians can read and compare the images without knowing which device was used to take the images and without being concerned about the sources of the

images. The universal OCT browser can further be equipped with many image-processing tools, such as retinal segmentation algorithms and image enhancement methods. The browser can also allow interactions with the users, such as letting users choose the way they want to visualize the OCT image, for example, in 2D, 3D, as a horizontal or vertical cross-sectional image, in C-mode, extracting individual retinal layer's information, or view the entire retina as a whole. This will provide more information about the retinal tissues and enable browsing and analyzing OCT images in a more convenient fashion. Last but not least, with a universal OCT browser, using the same analysis method to process and analyze OCT images regardless of manufacturer will become feasible. As various analysis methods have been recognized as one of the major components resulting in measurement discrepancies, using the same analysis method to process normalized OCT data will definitely reduce the variability of quantitative measurements. The presented signal normalization method can be a core foundation of such an effort.

We have mentioned two major factors causing systematic measurement differences, one is using different OCT devices, and the other is using devices from the same manufacturer but with different iterations. With the assumption that the inter-device measurement differences are larger than intra-device differences, we only investigated how signal normalization minimizes the systematic measurement differences among different SD-OCT devices. It would be interesting to test how the proposed signal normalization minimizes the measurement differences between devices from the same manufacturer with different generations. Second, in the clinical validation, we tested the effects between Cirrus and RTVue and between Cirrus and Spectralis separately. It is helpful to compare multiple SD-OCT devices altogether. Toward the end of this study, we were lucky to get phantom eyes from U.S. Food and Drug Administration (FDA).[119, 120] Phantom eyes are designed and fabricated with scattering materials in a layered structure to

mimic human retina.[119, 120] As phantom eyes are more robust and do not change over time, we are able to use phantom eyes to test our signal normalization method in a more systematic way, such as how the OCT signal characteristics change in different conditions, or how the speckle noise affects the signals. Last, we would like to borrow the idea of signal morphing from audio signal processing.[121-123] Audio signals can be processed so that people speak with different frequencies, volumes, and pitches but end up with a similar sound. It would be interesting to see how signal morphing can help reduce the variation in OCT signal characteristics.

In summary, a novel step-wise signal normalization method was proposed in this study, which successfully reduces systematic differences and enables direct comparison among various OCT devices. The method will become a useful tool for OCT data normalization or standardization when multiple OCT devices are involved in the same study, and will broaden the use of OCT in both clinical and research applications to a more exciting field.

APPENDIX A

SUBJECTIVE EVALUATION QUESTIONNAIRE FOR QUALITATIVE ASSESSMENT

Signal Normalization Among Multiple Optical Coherence Tomography Devices

The purpose of this research study is to validate the effect of a self-developed signal normalization method on reducing the variation of image appearance among various spectral-domain optical coherence tomography (SD-OCT) devices.

In this questionnaire, 35 sets of images (5 sets in the Example session and 30 sets in the actual questionnaire) will be presented. Each set contains 2 SD-OCT cross-sectional images extracted from the same location acquired from the same eye at the same visit with any combination of 3 SD-OCT devices (Cirrus (Zeiss, Dublin, California, USA), RTVue (Optovue, Fremont, California, USA), and Spectralis (Heidelberg Engineering, Heidelberg, Germany)). There are healthy, glaucomatous, and retinal pathology cases. You will be asked to judge the relative similarity of image appearance for each set based on 1) the contrast between retinal signal and background noise, 2) the contrast between adjacent retinal layers, and 3) the textures of the retina. Please ignore the differences because of translation or rotation issue (overall shape of the retina) and ignore any pathologic context.

The similarity will be recorded in 5 levels: 1) Significantly different (0-20% similarity), 2) Somewhat different (20-40% similarity), 3) Cannot decide if it is similar (40-60% similarity), 4) Looks similar (60-80% similarity), and 5) Nearly identical (as if taken by the same device) (80-100% similarity)
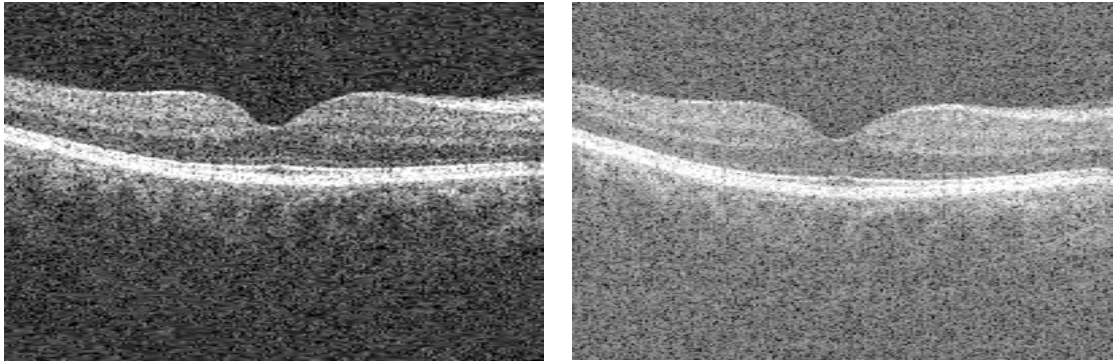
Please check (X) the description that best matches your visual experience / feeling.

Thank you very much for your participation!

Examples

The survey starts with some examples. The examples show the variety of the testing data and are presented in the same way as the actual questionnaire.
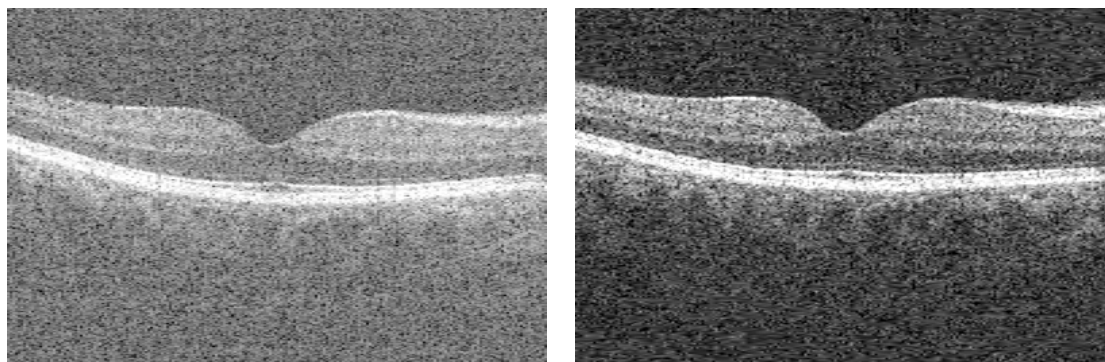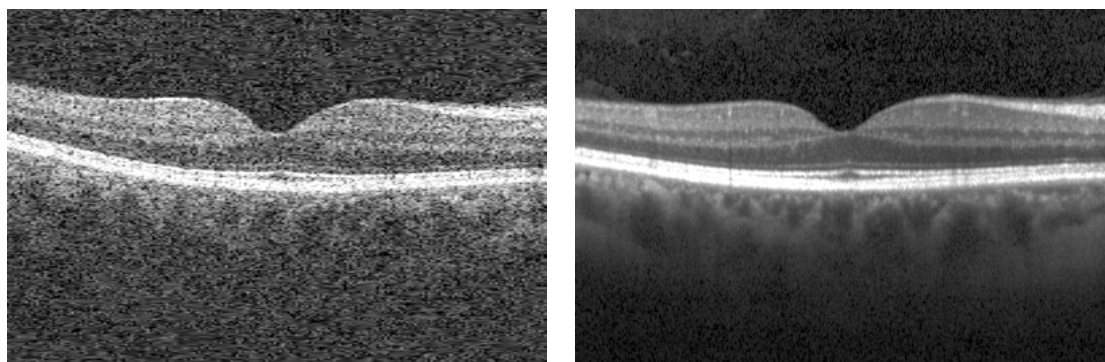
Example 1

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)
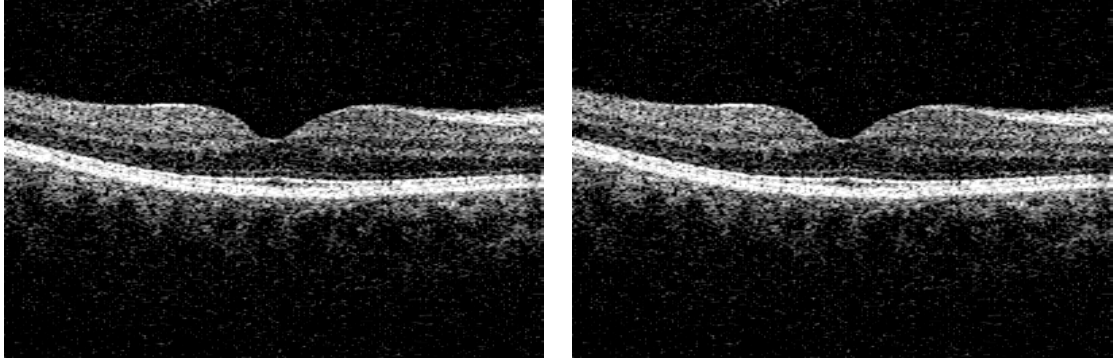
Example 2

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

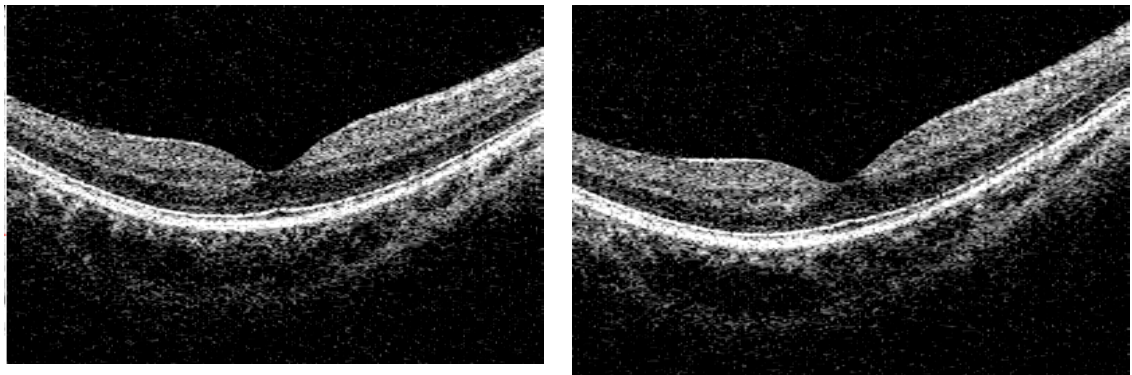( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

Example 3

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)
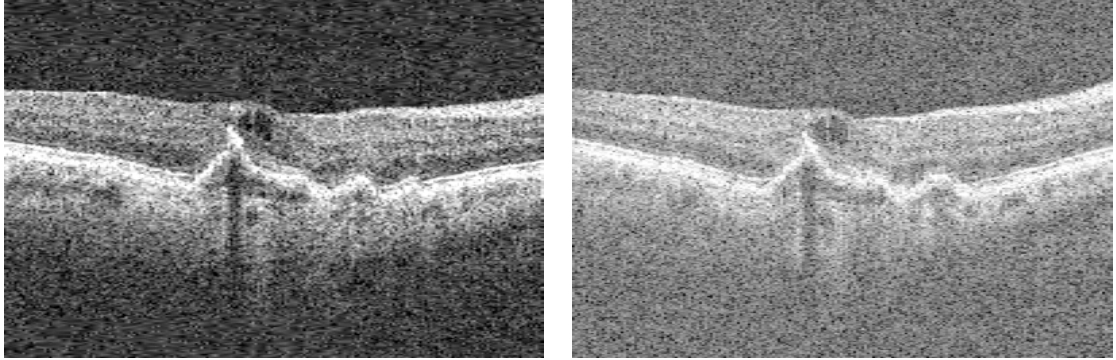
Example 4

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

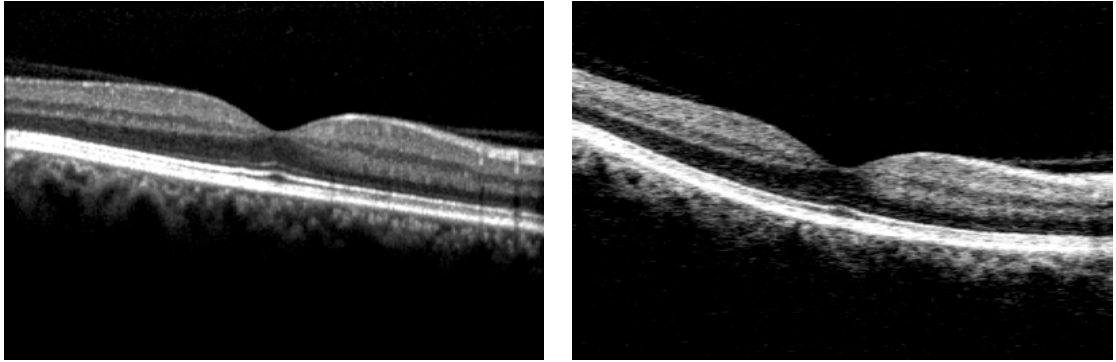( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

Example 5

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

Signal Normalization Method Among Multiple Optical Coherence Tomography Devices

Subjective Evaluation Questionnaire Starts

1.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)
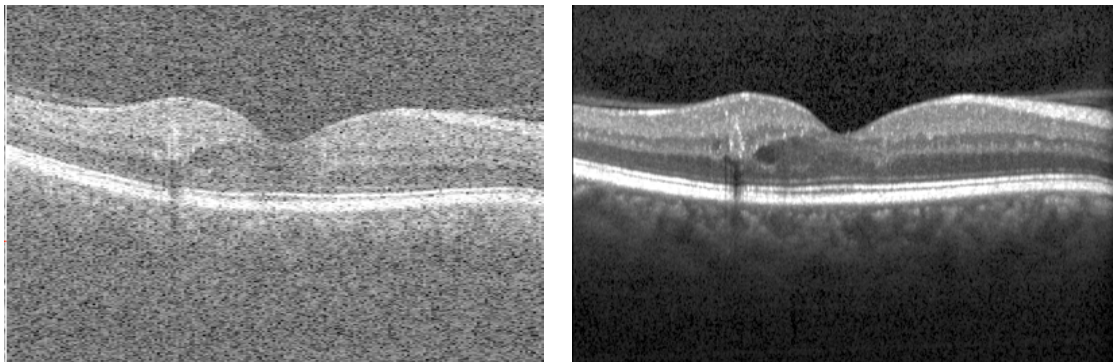
2.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

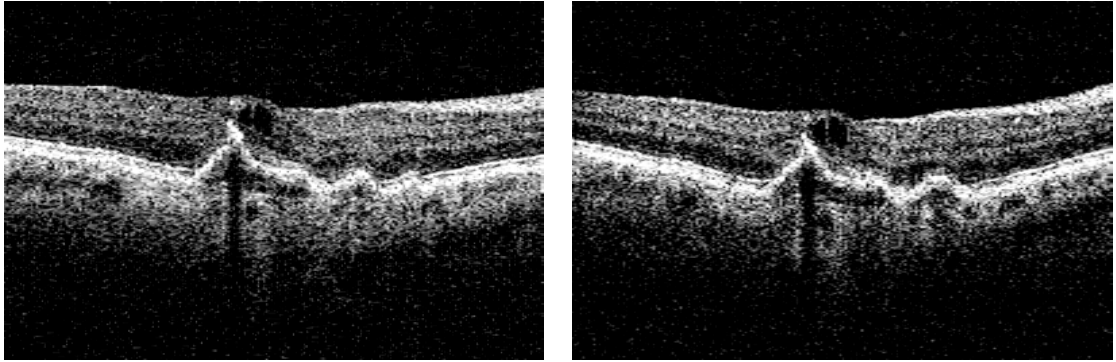( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

3.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

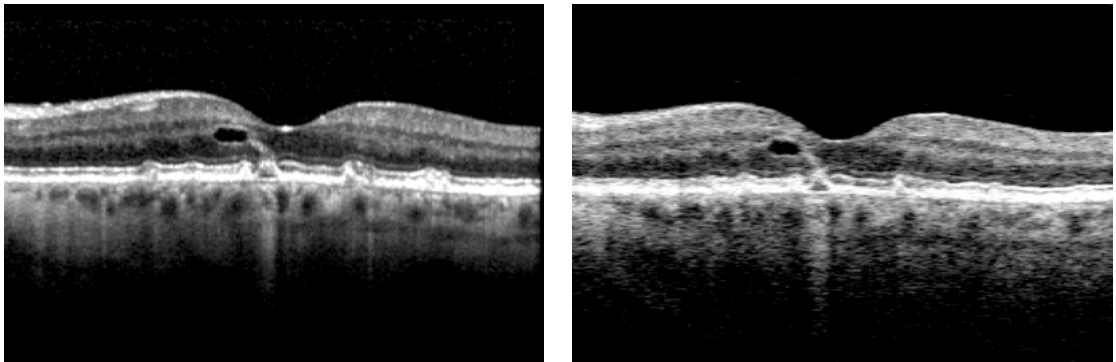( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

4.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

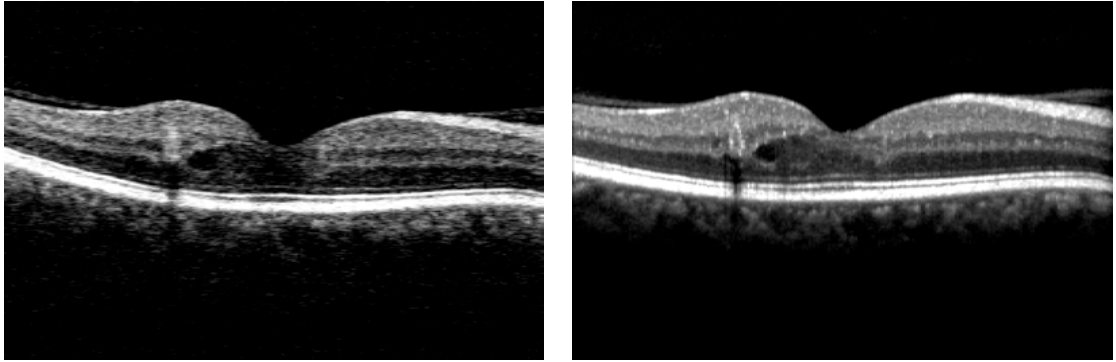( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

5.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

6.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

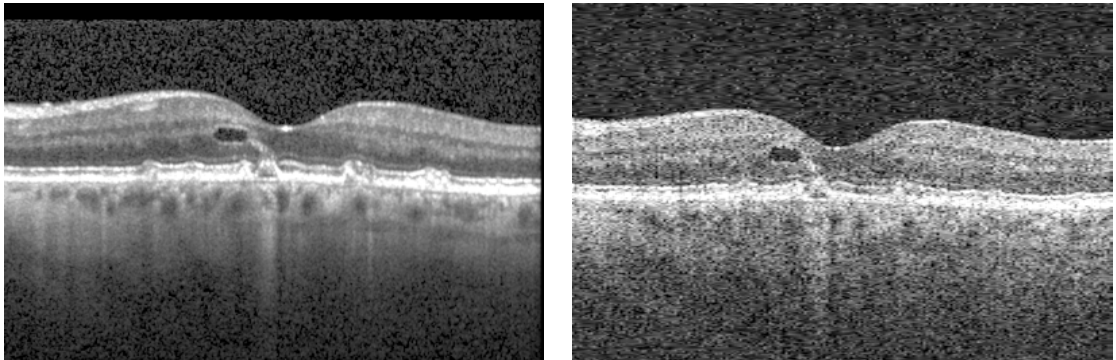( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

7.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

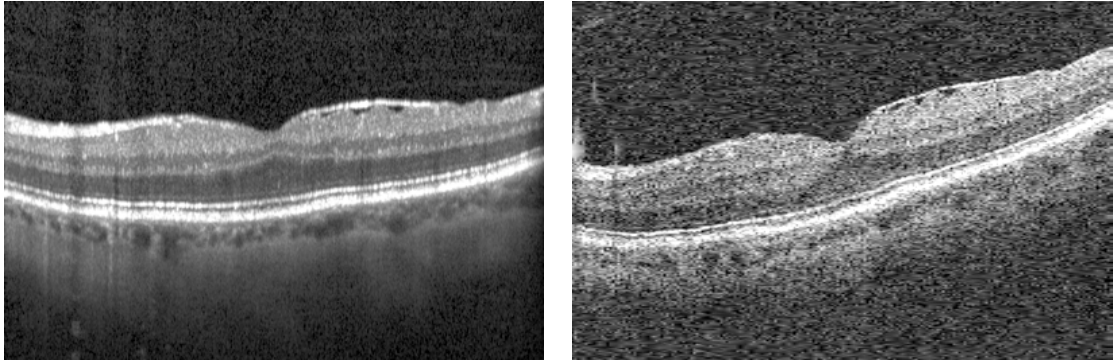( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

8.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

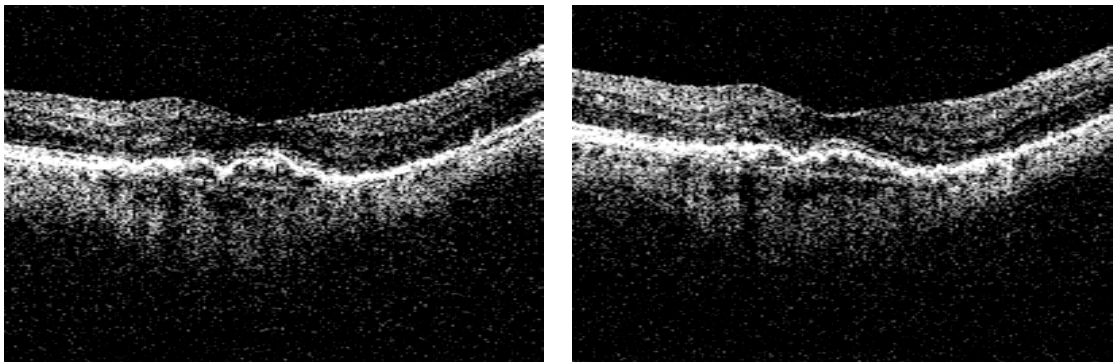( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

9.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

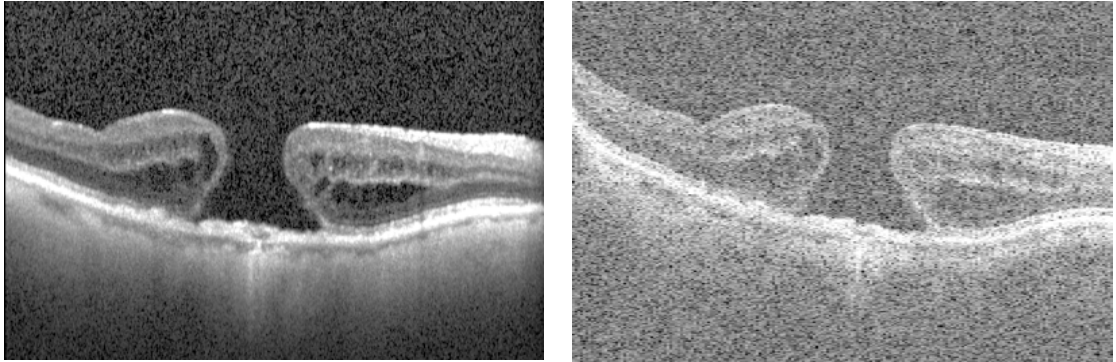( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

10.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

11.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

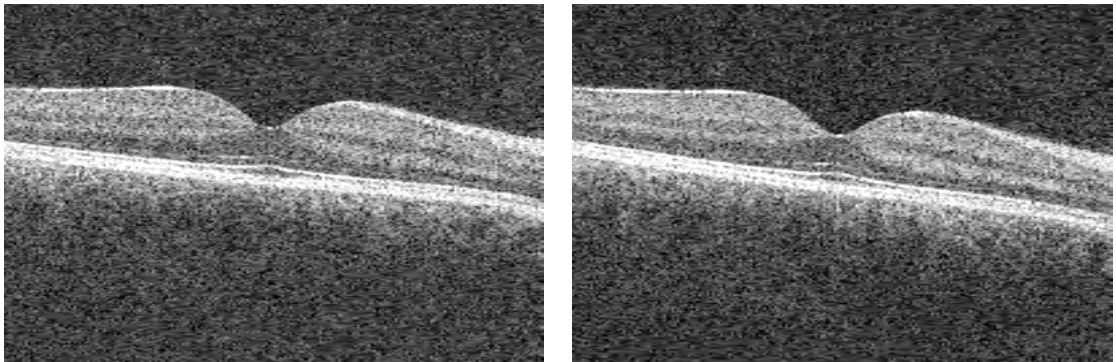( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

12.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

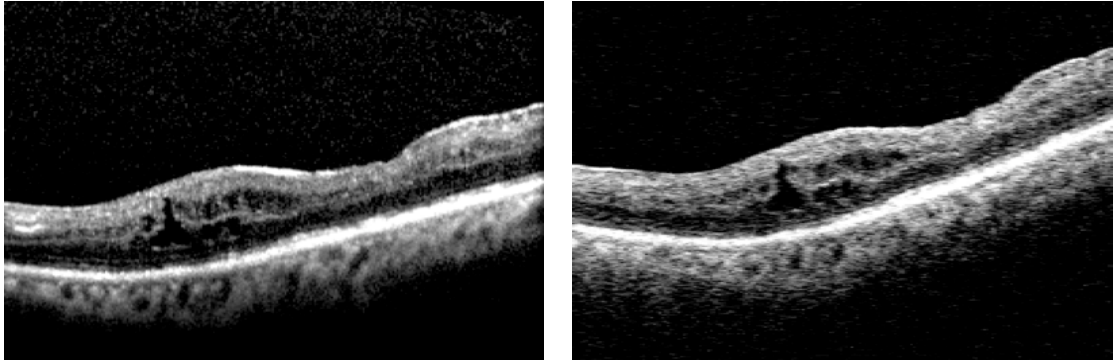( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

13.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

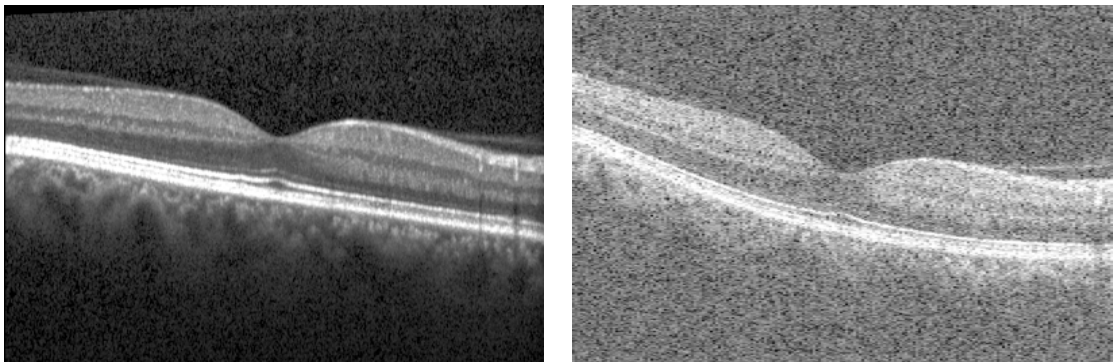( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

14.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

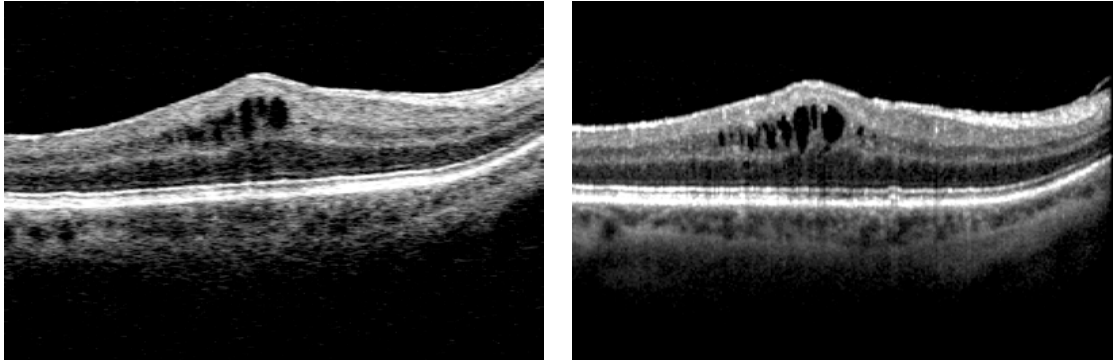( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

15.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)
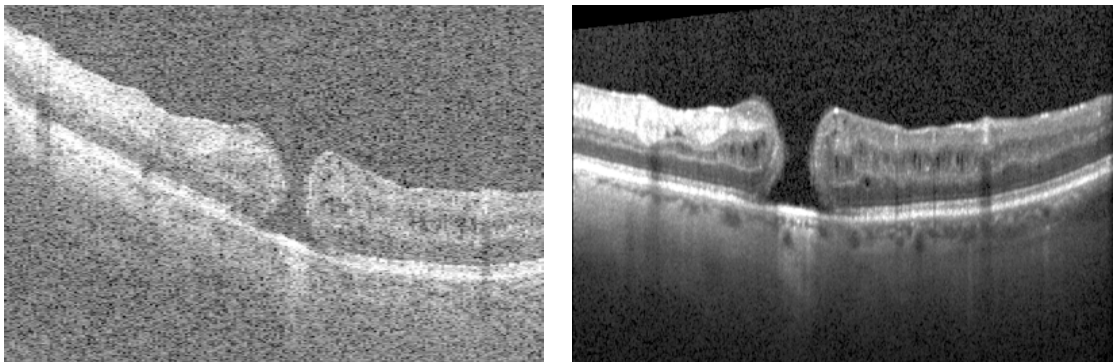
16.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

17.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

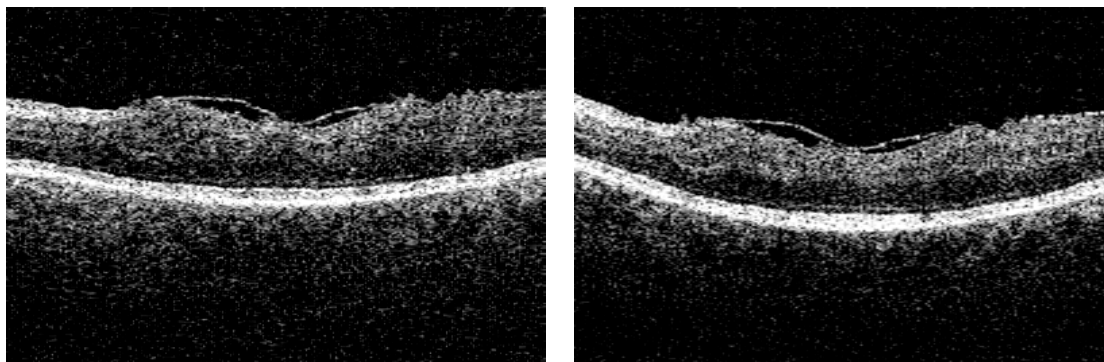( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

18.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

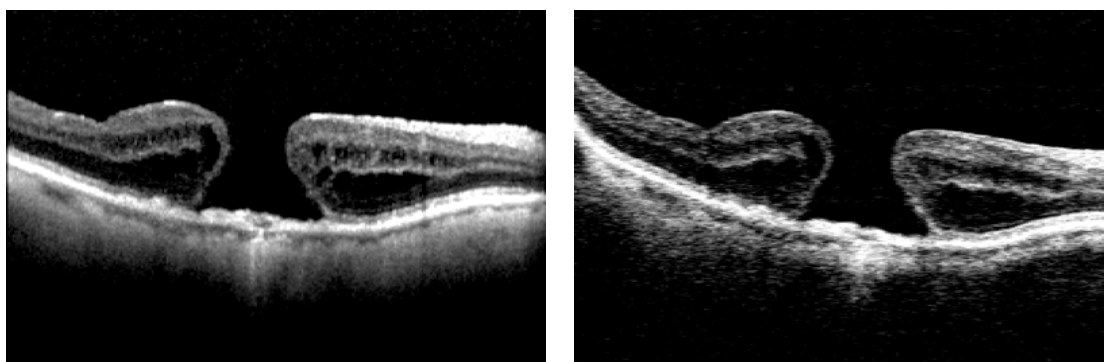( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

19.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

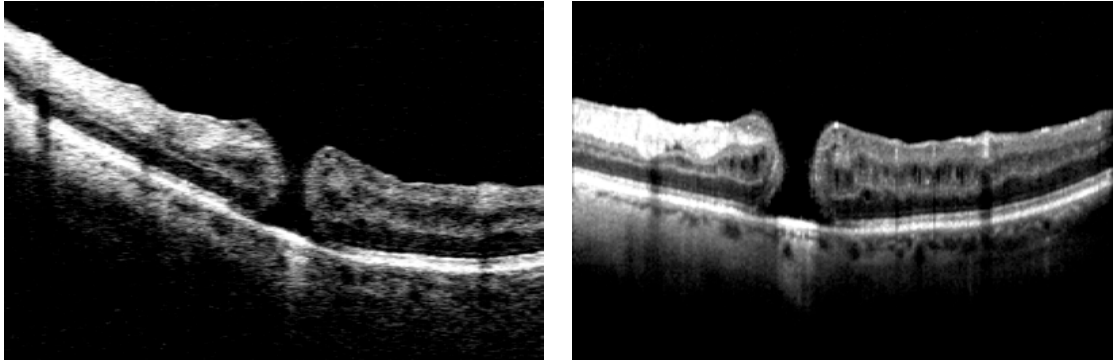( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

20.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

21.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

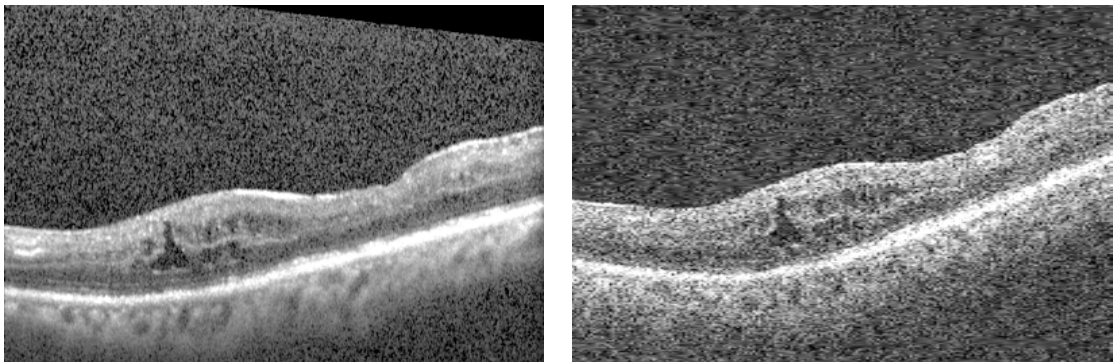( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

22.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

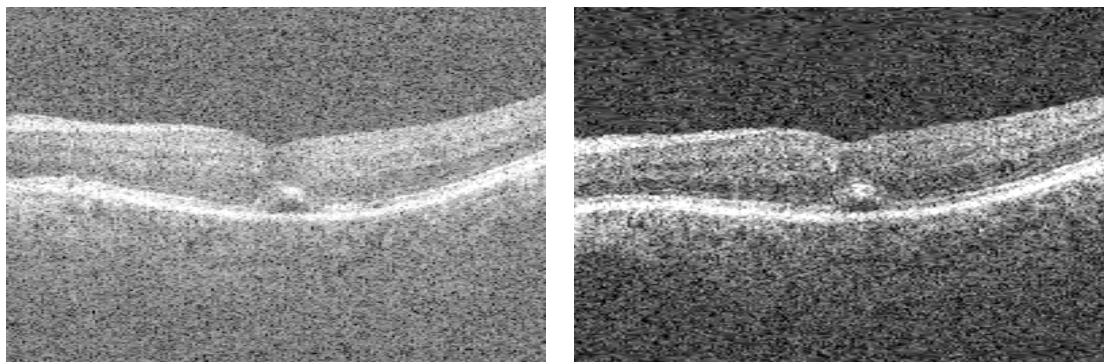( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

23.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

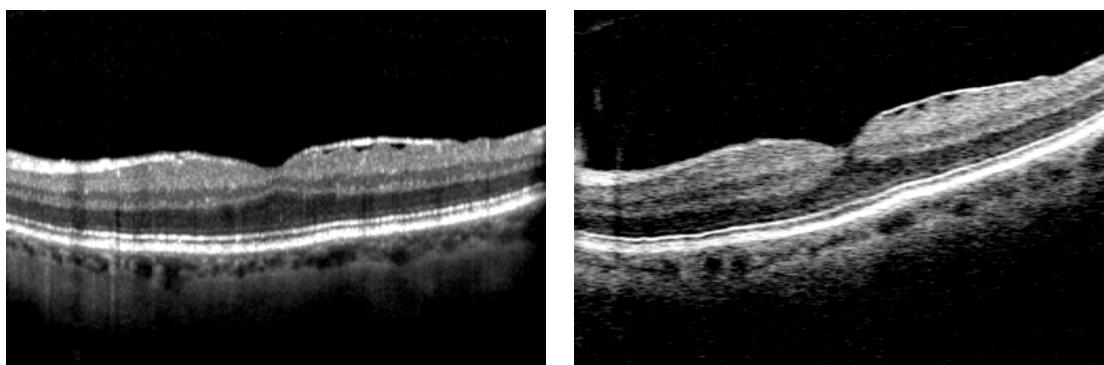( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

24.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

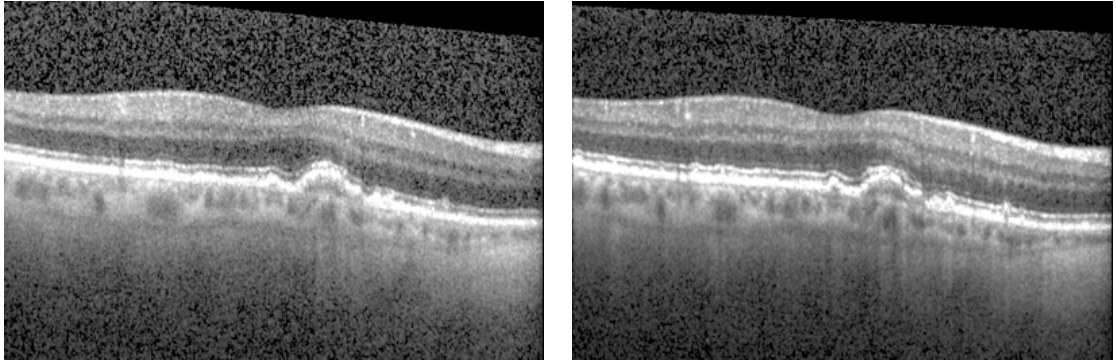( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

25.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)
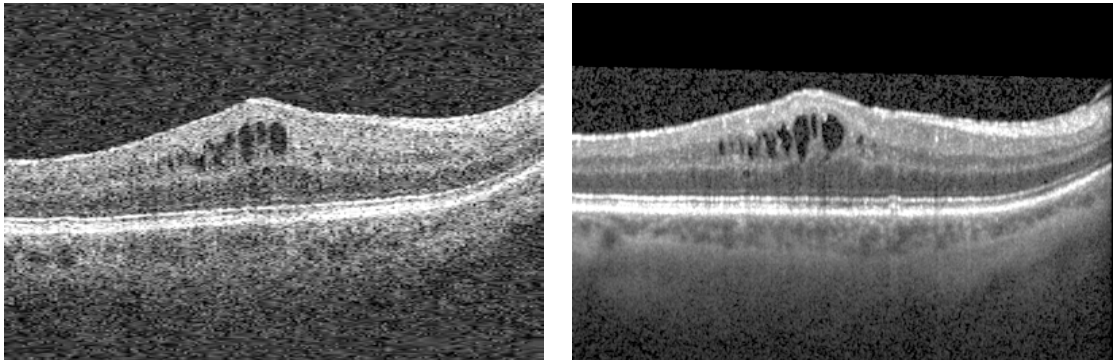
26.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

27.
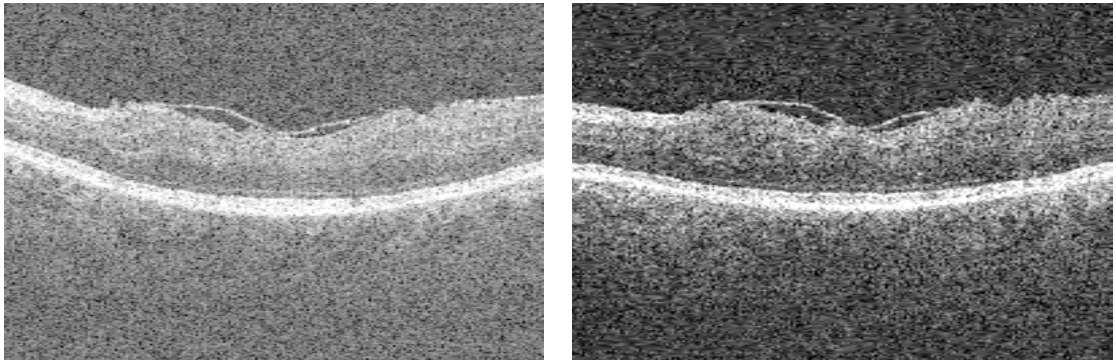
Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

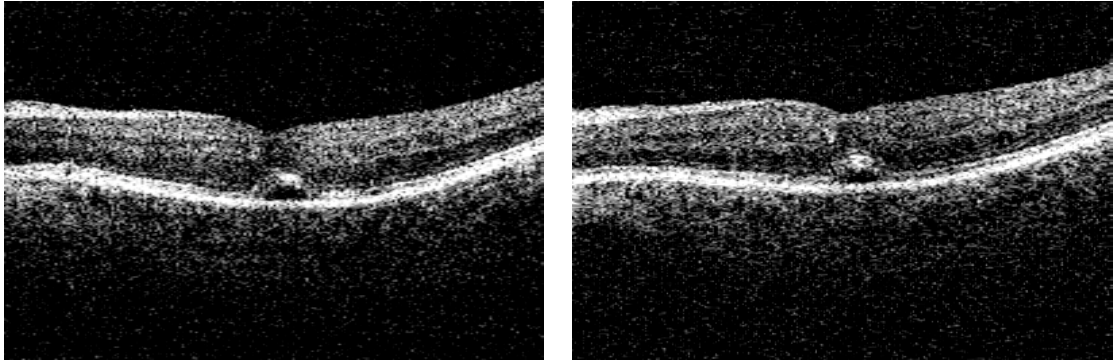( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

28.
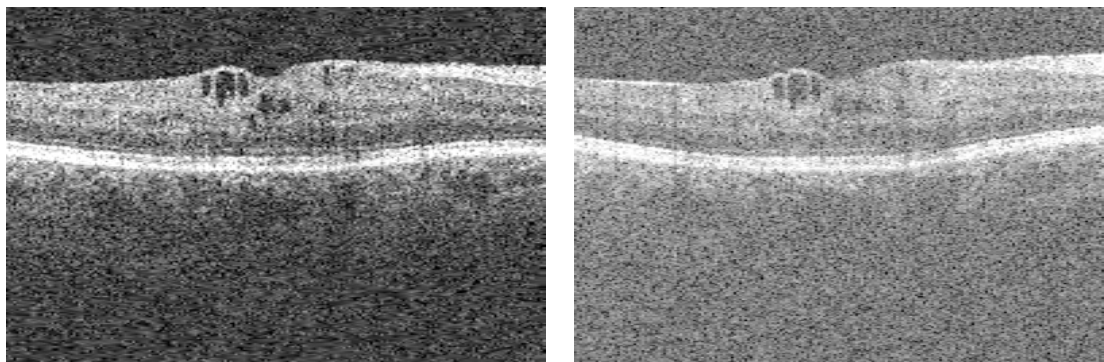
Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

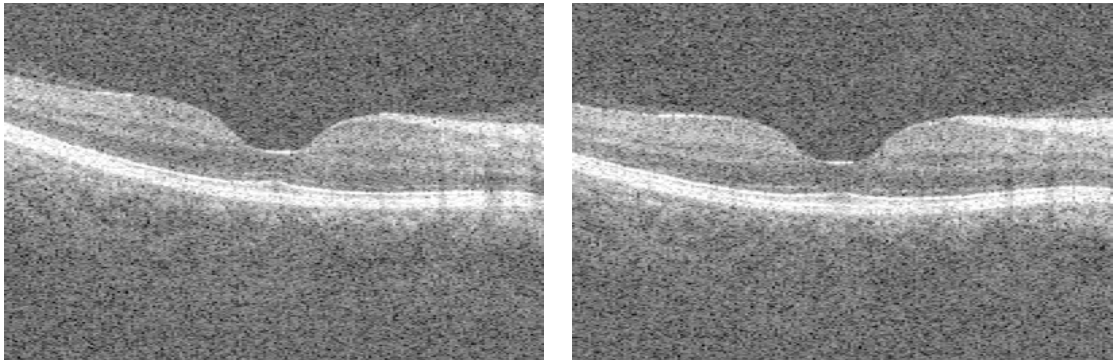( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

29.
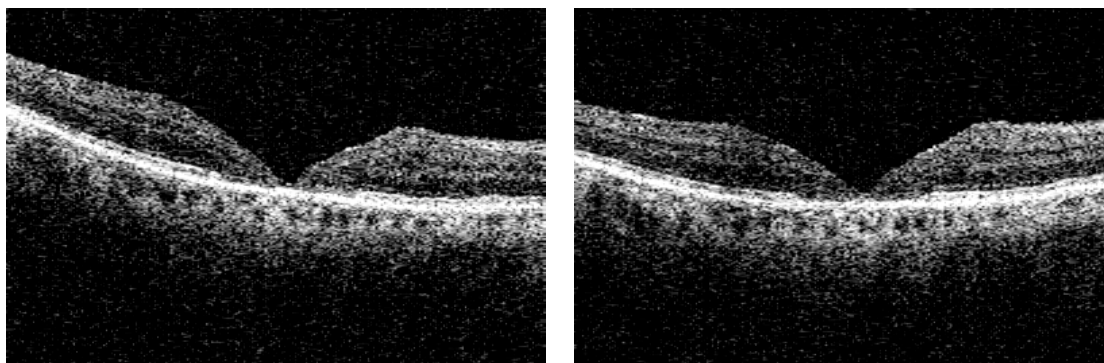
Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

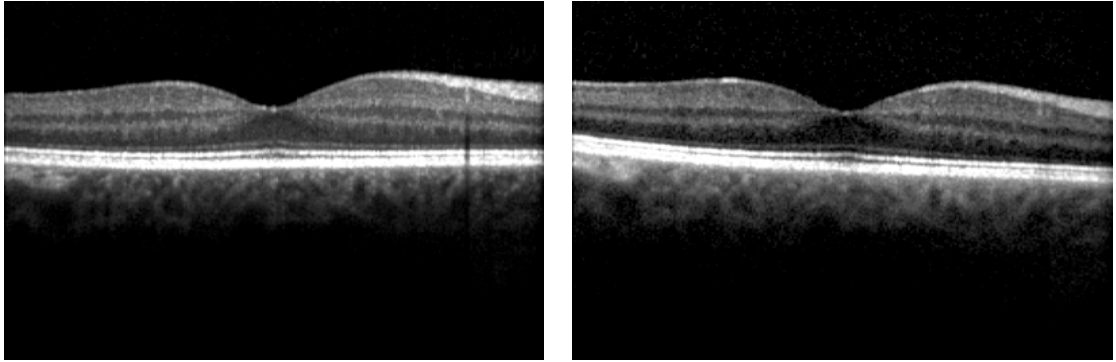( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)

30.

Please judge the similarity between the two images in terms of intensity, contrast, and overall image appearance based on your visual experience:

( ) 1 Significantly different (0-20% similarity)

( ) 2 Somewhat different (20-40% similarity)

( ) 3 Cannot decide if it is similar (40-60% similarity)

( ) 4 Looks similar (60-80% similarity)

( ) 5 Nearly identical (as if taken by the same device) (80-100% similarity)
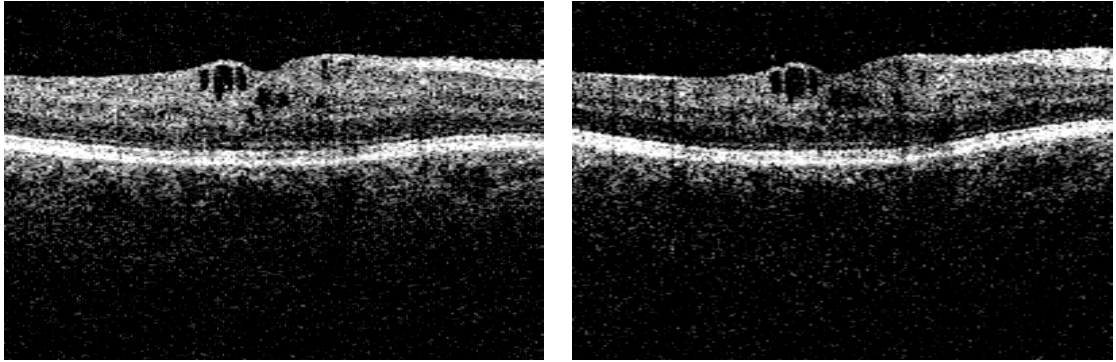
BIBLIOGRAPHY

1. Huang D, Swanson E, Lin C, Schuman J, Stinson W, Chang W, Hee M, Flotte T, Gregory K, Puliafito C, et a. Optical coherence tomography. *Science*. 1991;254(5035):1178-1181.

2. Drexler W, Fujimoto JG. State-of-the-art retinal optical coherence tomography. *Prog Retin Eye Res*. 2008;27(1):45-88.

3. Schuman JS, Puliafito CA, Fujimoto JG, Duker JS. Optical Coherence Tomography of Ocular Diseases: SLACK Incorporated; 2013.

4. Fercher AF, Hitzenberger CK, Kamp G, El-Zaiat SY. Measurement of intraocular distances by backscattering spectral interferometry. *Opt Commun*. 1995;117(1-2):43-48.

5. Gabriele ML, Wollstein G, Ishikawa H, Xu J, Kim J, Kagemann L, Folio LS, Schuman JS. Three dimensional optical coherence tomography imaging: advantages and advances. *Prog Retin Eye Res*. 2010;29(6):556-579.

6. Swanson EA, Huang D. Ophthalmic OCT Reaches $1 Billion Per Year: But reimbursement clampdown clouds future innovation. Retinal Physician. 2011:45, 58-59, 62.

7. Born M, Wolf E, Bhatia AB, Clemmow PC, Gabor D, Stokes AR, Taylor AM, Wayman PA, Wilcock WL. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light: Cambridge University Press; 1999.

8. Youngquist RC, Carr S, Davies DE. Optical coherence-domain reflectometry: a new optical evaluation technique. *Opt Lett*. 1987;12(3):158-160.

9. 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(9):2699-2720.

10. Wojtkowski M, Srinivasan V, Ko T, Fujimoto J, Kowalczyk A, Duker J. Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation. *Opt Express*. 2004;12(11):2404-2422.

11.   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(3):457-463.

12.   Nassif N, Cense B, Park BH, Yun SH, Chen TC, Bouma BE, Tearney GJ, de Boer JF. In vivo human retinal imaging by ultrahigh-speed spectral domain optical coherence tomography. *Opt Lett*. 2004;29(5):480-482.

13.   Cense B, Nassif N, Chen T, Pierce M, Yun SH, Park B, Bouma B, Tearney G, de Boer J. Ultrahigh-resolution high-speed retinal imaging using spectral-domain optical coherence tomography. *Opt Express*. 2004;12(11):2435-2447.

14.   Leitgeb R, Hitzenberger C, Fercher A. Performance of fourier domain vs. time domain optical coherence tomography. *Opt Express*. 2003;11(8):889-894.

15.   de Boer JF, Cense B, Park BH, Pierce MC, Tearney GJ, Bouma BE. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. *Opt Lett*. 2003;28(21):2067-2069.

16.   Choma M, Sarunic M, Yang C, Izatt J. Sensitivity advantage of swept source and Fourier domain optical coherence tomography. *Opt Express*. 2003;11(18):2183-2189.

17.   Grulkowski I, Liu JJ, Potsaid B, Jayaraman V, Lu CD, Jiang J, Cable AE, Duker JS, Fujimoto JG. Retinal, anterior segment and full eye imaging using ultrahigh speed swept source OCT with vertical-cavity surface emitting lasers. *Biomed Opt Express*. 2012;3(11):2733-2751.

18.   Leung CK. Diagnosing glaucoma progression with optical coherence tomography. *Curr Opin Ophthalmol*. 2014;25(2):104-111.

19.   Alam S, Zawadzki RJ, Choi S, Gerth C, Park SS, Morse L, Werner JS. Clinical application of rapid serial fourier-domain optical coherence tomography for macular imaging. *Ophthalmology*. 2006;113(8):1425-1431.

20.   Chen TC, Cense B, Pierce MC, Nassif N, Park BH, Yun SH, White BR, Bouma BE, Tearney GJ, de Boer JF. Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging. *Arch Ophthalmol*. 2005;123(12):1715-1720.

21.   Nassif N, Cense B, Park B, Pierce M, Yun S, Bouma B, Tearney G, Chen T, de Boer J. In vivo high-resolution video-rate spectral-domain optical coherence tomography of the human retina and optic nerve. *Opt Express*. 2004;12(3):367-376.

22.   Buchser NM, Wollstein G, Ishikawa H, Bilonick RA, Ling Y, Folio LS, Kagemann L, Noecker RJ, Albeiruti E, Schuman JS. 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(7):3742-3747.

23. Pavlin CJ, Harasiewicz K, Sherar MD, Foster FS. Clinical use of ultrasound biomicroscopy. *Ophthalmology*. 1991;98(3):287-295.

24. Baumal CR. Clinical applications of optical coherence tomography. *Curr Opin Ophthalmol*. 1999;10(3):182-188.

25. Costa RA, Skaf M, Melo LA, Jr., Calucci D, Cardillo JA, Castro JC, Huang D, Wojtkowski M. Retinal assessment using optical coherence tomography. *Prog Retin Eye Res*. 2006;25(3):325-353.

26. Sakata LM, Deleon-Ortega J, Sakata V, Girkin CA. Optical coherence tomography of the retina and optic nerve - a review. *Clin Experiment Ophthalmol*. 2009;37(1):90-99.

27. Williams ZY, Schuman JS, Gamell L, Nemi A, Hertzmark E, Fujimoto JG, Mattox C, Simpson J, Wollstein G. Optical coherence tomography measurement of nerve fiber layer thickness and the likelihood of a visual field defect. *Am J Ophthalmol*. 2002;134(4):538-546.

28. Schuman JS, Hee MR, Puliafito CA, Wong C, Pedut-Kloizman T, Lin CP, Hertzmark E, Izatt JA, Swanson EA, Fujimoto JG. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. *Arch Ophthalmol*. 1995;113(5):586-596.

29. Deleon-Ortega JE, Arthur SN, McGwin G, Jr., Xie A, Monheit BE, Girkin CA. Discrimination between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective optic nerve head assessment. *Invest Ophthalmol Vis Sci*. 2006;47(8):3374-3380.

30. Puliafito CA, Hee MR, Lin CP, Reichel E, Schuman JS, Duker JS, Izatt JA, Swanson EA, Fujimoto JG. Imaging of macular diseases with optical coherence tomography. *Ophthalmology*. 1995;102(2):217-229.

31. Chan A, Duker JS, Schuman JS, Fujimoto JG. Stage 0 macular holes: observations by optical coherence tomography. *Ophthalmology*. 2004;111(11):2027-2032.

32. Johnson MW, Van Newkirk MR, Meyer KA. Perifoveal vitreous detachment is the primary pathogenic event in idiopathic macular hole formation. *Arch Ophthalmol*. 2001;119(2):215-222.

33. Johnson MW. Improvements in the understanding and treatment of macular hole. *Curr Opin Ophthalmol*. 2002;13(3):152-160.

34. van Velthoven ME, Faber DJ, Verbraak FD, van Leeuwen TG, de Smet MD. Recent developments in optical coherence tomography for imaging the retina. *Prog Retin Eye Res*. 2007;26(1):57-77.

35. Brown DM, Regillo CD. Anti-VEGF agents in the treatment of neovascular age-related macular degeneration: applying clinical trial results to the treatment of everyday patients. *Am J Ophthalmol*. 2007;144(4):627-637.

36. Jaffe GJ, Caprioli J. Optical coherence tomography to detect and manage retinal disease and glaucoma. *Am J Ophthalmol*. 2004;137(1):156-169.

37. Quigley HA, Dunkelberger GR, Green WR. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *Am J Ophthalmol*. 1989;107(5):453-464.

38. Quigley HA, Addicks EM, Green WR. Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fiber loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. *Arch Ophthalmol*. 1982;100(1):135-146.

39. Hsu SY, Tsai RK. Analysis of retinal nerve fiber layer and macular thickness measurements in healthy Taiwanese individuals using optical coherence tomography (Stratus OCT). *J Glaucoma*. 2008;17(1):30-35.

40. DeLeon Ortega JE, Sakata LM, Kakati B, McGwin G, Jr., Monheit BE, Arthur SN, Girkin CA. Effect of glaucomatous damage on repeatability of confocal scanning laser ophthalmoscope, scanning laser polarimetry, and optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2007;48(3):1156-1163.

41. Carpineto P, Ciancaglini M, Aharrh-Gnama A, Cirone D, Mastropasqua L. Custom measurement of retinal nerve fiber layer thickness using STRATUS OCT in normal eyes. *Eur J Ophthalmol*. 2005;15(3):360-366.

42. Budenz DL, Fredette MJ, Feuer WJ, Anderson DR. Reproducibility of peripapillary retinal nerve fiber thickness measurements with stratus OCT in glaucomatous eyes. *Ophthalmology*. 2008;115(4):661-666.e664.

43. Zangwill LM, Bowd C, Berry CC, Williams J, Blumenthal EZ, Sanchez-Galeana CA, Vasile C, Weinreb RN. Discriminating between normal and glaucomatous eyes using the Heidelberg Retina Tomograph, GDx Nerve Fiber Analyzer, and Optical Coherence Tomograph. *Arch Ophthalmol*. 2001;119(7):985-993.

44. Medeiros FA, Zangwill LM, Bowd C, Vessani RM, Susanna R, Jr., Weinreb RN. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. *Am J Ophthalmol*. 2005;139(1):44-55.

45. Kanamori A, Nagai-Kusuhara A, Escano MF, Maeda H, Nakamura M, Negi A. Comparison of confocal scanning laser ophthalmoscopy, scanning laser polarimetry and optical coherence tomography to discriminate ocular hypertension and glaucoma at an early stage. *Graefes Arch Clin Exp Ophthalmol*. 2006;244(1):58-68.

46. Greaney MJ, Hoffman DC, Garway-Heath DF, Nakla M, Coleman AL, Caprioli J. Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma. *Invest Ophthalmol Vis Sci*. 2002;43(1):140-145.

47. Bowd C, Weinreb RN, Williams JM, Zangwill LM. The retinal nerve fiber layer thickness in ocular hypertensive, normal, and glaucomatous eyes with optical coherence tomography. *Arch Ophthalmol*. 2000;118(1):22-26.

48. Wollstein G, Schuman JS, Price LL, Aydin A, Stark PC, Hertzmark E, Lai E, Ishikawa H, Mattox C, Fujimoto JG, Paunescu LA. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol*. 2005;123(4):464-470.

49. Medeiros FA, Zangwill LM, Alencar LM, Bowd C, Sample PA, Susanna R, Jr., Weinreb RN. Detection of glaucoma progression with stratus OCT retinal nerve fiber layer, optic nerve head, and macular thickness measurements. *Invest Ophthalmol Vis Sci*. 2009;50(12):5741-5748.

50. Maeda N. Optical coherence tomography for corneal diseases. *Eye Contact Lens*. 2010;36(5):254-259.

51. Simpson T, Fonn D. Optical coherence tomography of the anterior segment. *Ocul Surf*. 2008;6(3):117-127.

52. Ramos JL, Li Y, Huang D. Clinical and research applications of anterior segment optical coherence tomography - a review. *Clin Experiment Ophthalmol*. 2009;37(1):81-89.

53. Konstantopoulos A, Hossain P, Anderson DF. Recent advances in ophthalmic anterior segment imaging: a new era for ophthalmic diagnosis? *Br J Ophthalmol*. 2007;91(4):551-557.

54. Chen J, Lee L. Clinical applications and new developments of optical coherence tomography: an evidence-based review. *Clin Exp Optom*. 2007;90(5):317-335.

55. Gabriele ML, Wollstein G, Ishikawa H, Kagemann L, Xu J, Folio LS, Schuman JS. Optical coherence tomography: history, current status, and laboratory work. *Invest Ophthalmol Vis Sci*. 2011;52(5):2425-2436.

56. Seibold LK, Mandava N, Kahook MY. Comparison of retinal nerve fiber layer thickness in normal eyes using time-domain and spectral-domain optical coherence tomography. *Am J Ophthalmol*. 2010;150(6):807-814.

57. Giani A, Cigada M, Choudhry N, Deiro AP, Oldani M, Pellegrini M, Invernizzi A, Duca P, Miller JW, Staurenghi G. Reproducibility of retinal thickness measurements on normal and pathologic eyes by different optical coherence tomography instruments. *Am J Ophthalmol*. 2010;150(6):815-824.

58.   Leite MT, Rao HL, Weinreb RN, Zangwill LM, Bowd C, Sample PA, Tafreshi A, Medeiros FA. Agreement among spectral-domain optical coherence tomography instruments for assessing retinal nerve fiber layer thickness. *Am J Ophthalmol*. 2011;151(1):85-92.e81.

59.   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(6):832-837.

60.   Wollstein G, Schuman JS, Price LL, Aydin A, Beaton SA, Stark PC, Fujimoto JG, Ishikawa H. Optical coherence tomography (OCT) macular and peripapillary retinal nerve fiber layer measurements and automated visual fields. *Am J Ophthalmol*. 2004;138(2):218-225.

61.   DICOM standards committee. Proposal for collaboration between DICOM and the International Working Group for intracoronary OCT standardization and validation (IWG-OCT). http://medical.nema.org/dicom/minutes/committee/2009/2009-04-21/other_documents/wg-01_proposed collaboration_iwg-oct.doc Accessed September 4, 2012.

62.   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(4):1280-1285.

63.   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(1):16-22.

64.   Vizzeri G, Weinreb RN, Gonzalez-Garcia AO, Bowd C, Medeiros FA, Sample PA, Zangwill LM. Agreement between spectral-domain and time-domain OCT for measuring RNFL thickness. *Br J Ophthalmol*. 2009;93(6):775-781.

65.   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(7):1264-1270, 1270.e1261.

66.   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(7):1271-1277.

67.   Wu Z, Vazeen M, Varma R, Chopra V, Walsh AC, LaBree LD, Sadda SR. Factors associated with variability in retinal nerve fiber layer thickness measurements obtained by optical coherence tomography. *Ophthalmology*. 2007;114(8):1505-1512.

68.   Cheung CY, Leung CK, Lin D, Pang CP, Lam DS. Relationship between retinal nerve fiber layer measurement and signal strength in optical coherence tomography. *Ophthalmology*. 2008;115(8):1347-1351, 1351.e1341-1342.

69.   van Velthoven ME, van der Linden MH, de Smet MD, Faber DJ, Verbraak FD. Influence of cataract on optical coherence tomography image quality and retinal thickness. *Br J Ophthalmol*. 2006;90(10):1259-1262.

70.   Stein DM, Wollstein G, Ishikawa H, Hertzmark E, Noecker RJ, Schuman JS. Effect of corneal drying on optical coherence tomography. *Ophthalmology*. 2006;113(6):985-991.

71.   Kok PH, van Dijk HW, van den Berg TJ, Verbraak FD. A model for the effect of disturbances in the optical media on the OCT image quality. *Invest Ophthalmol Vis Sci*. 2009;50(2):787-792.

72.   Barkana Y, Burgansky-Eliash Z, Gerber Y, Melamed S, Neudorfer M, Avni I, Bartov E, Morad Y. Inter-device variability of the Stratus optical coherence tomography. *Am J Ophthalmol*. 2009;147(2):260-266.

73.   Kim JS, Ishikawa H, Gabriele ML, Wollstein G, Bilonick RA, Kagemann L, Fujimoto JG, Schuman JS. Retinal nerve fiber layer thickness measurement comparability between time domain optical coherence tomography (OCT) and spectral domain OCT. *Invest Ophthalmol Vis Sci*. 2010;51(2):896-902.

74.   van der Schoot J, Vermeer KA, de Boer JF, Lemij HG. The Effect of Glaucoma on the Optical Attenuation Coefficient of the Retinal Nerve Fiber Layer in Spectral Domain Optical Coherence Tomography Images. *Invest Ophthalmol Vis Sci*. 2012;53(4):2424-2430.

75.   Goodman JW. Some fundamental properties of speckle. *JOSA*. 1976;66(11):1145-1150.

76.   Schmitt JM, Xiang S, Yung KM. Speckle in optical coherence tomography. *Journal of Biomedical Optics*. 1999;4(1):95-105.

77.   Ramrath L, Moreno G, Mueller H, Bonin T, Huettmann G, Schweikard A. Towards multi-directional OCT for speckle noise reduction. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008: Springer; 2008. p. 815-823.

78.   Li Z, Li H, He Y, Cai S, Xie S. A model of speckle contrast in optical coherence tomography for characterizing the scattering coefficient of homogenous tissues. *Physics in medicine and biology*. 2008;53(20):5859.

79.   Chitchian S, Fiddy M, Fried NM. Wavelet denoising during optical coherence tomography of the prostate nerves using the complex wavelet transform. *Engineering in Medicine and Biology Society, 2008 EMBS 2008 30th Annual International Conference of the IEEE*; 2008: IEEE.

80. Puvanathasan P, Bizheva K. Interval type-II fuzzy anisotropic diffusion algorithm for speckle noise reduction in optical coherence tomography images. *Optics express*. 2009;17(2):733-746.

81. Mishra A, Wong A, Bizheva K, Clausi DA. Intra-retinal layer segmentation in optical coherence tomography images. *Optics express*. 2009;17(26):23719-23728.

82. Adler DC, Ko TH, Fujimoto JG. Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter. *Optics letters*. 2004;29(24):2878-2880.

83. Puvanathasan P, Bizheva K. Speckle noise reduction algorithm for optical coherence tomography based on interval type II fuzzy set. *Optics express*. 2007;15(24):15747-15758.

84. Ozcan A, Bilenca A, Desjardins AE, Bouma BE, Tearney GJ. Speckle reduction in optical coherence tomography images using digital filtering. *JOSA A*. 2007;24(7):1901-1910.

85. Ishikawa H, Stein DM, Wollstein G, Beaton S, Fujimoto JG, Schuman JS. Macular Segmentation with Optical Coherence Tomography. *Invest Ophthalmol Vis Sci*. 2005;46(6):2012-2017.

86. Sung KR, Wollstein G, Schuman JS, Bilonick RA, Ishikawa H, Townsend KA, Kagemann L, Gabriele ML, Group ftAIiGS. Scan quality effect on glaucoma discrimination by glaucoma imaging devices. *Br J Ophthalmol*. 2009;93(12):1580-1584.

87. Stein DM, Ishikawa H, Hariprasad R, Wollstein G, Noecker RJ, Fujimoto JG, Schuman JS. A new quality assessment parameter for optical coherence tomography. *British Journal of Ophthalmology*. 2006;90(2):186-190.

88. Huang Y, Gangaputra S, Lee KE, Narkar AR, Klein R, Klein BEK, Meuer SM, Danis RP. Signal Quality Assessment of Retinal Optical Coherence Tomography Images. *Invest Ophthalmol Vis Sci*. 2012;53(4):2133-2141.

89. Huang J, Liu X, Wu Z, Sadda S. Image quality affects macular and retinal nerve fiber layer thickness measurements on fourier-domain optical coherence tomography. *Ophthalmic Surg Lasers Imaging*. 2011;42(3):216-221.

90. Salinas HM, Fernandez DC. Comparison of PDE-based nonlinear diffusion approaches for image enhancement and denoising in optical coherence tomography. *IEEE Trans Med Imaging*. 2007;26(6):761-771.

91. Liu Y, Liang Y, Tong Z, Zhu X, Mu G. Contrast enhancement of optical coherence tomography images using least squares fitting and histogram matching. *Optics Communications*. 2007;279(1):23-26.

92. Adler D, Ko T, Herz P, Fujimoto J. Optical coherence tomography contrast enhancement using spectroscopic analysis with spectral autocorrelation. *Opt Express*. 2004;12(22):5487-5501.

93. Debevec PE, Malik J. Recovering high dynamic range radiance maps from photographs. ACM SIGGRAPH 2008 classes; Los Angeles, California. 1401174: ACM; 2008. p. 1-10.

94. Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans Graph*. 2002;21(3):257-266.

95. Reinhard E, Heidrich W, Debevec P, Pattanaik S, Ward G, Myszkowski K. High dynamic range imaging: acquisition, display, and image-based lighting: Morgan Kaufmann; 2010.

96. Bandoh Y, Qiu G, Okuda M, Daly S, Aach T, Au OC. Recent advances in high dynamic range imaging technology. *Image Processing (ICIP), 2010 17th IEEE International Conference on*; 2010: IEEE.

97. Fei P, Yu Z, Wang X, Lu PJ, Fu Y, He Z, Xiong J, Huang Y. High dynamic range optical projection tomography (HDR-OPT). *Opt Express*. 2012;20(8):8824-8836.

98. Dedrick E, Lau D. A Kalman-filtering approach to high dynamic range imaging for measurement applications. *Image Processing, IEEE Transactions on*. 2012;21(2):527-536.

99. Paunescu LA, Schuman JS, Price LL, Stark PC, Beaton S, Ishikawa H, Wollstein G, Fujimoto JG. Reproducibility of Nerve Fiber Thickness, Macular Thickness, and Optic Nerve Head Measurements Using StratusOCT. *Invest Ophthalmol Vis Sci*. 2004;45(6):1716-1724.

100. Sintorn IM, Bischof L, Jackway P, Haggarty S, Buckley M. Gradient based intensity normalization. *Journal of microscopy*. 2010;240(3):249-258.

101. Stanciu SG, Stanciu GA, Coltuc D. Automated compensation of light attenuation in confocal microscopy by exact histogram specification. *Microscopy research and technique*. 2010;73(3):165-175.

102. Salas-Gonzalez D, Gorriz JM, Ramirez J, Padilla P, Illan IA. Improving the convergence rate in affine registration of PET and SPECT brain images using histogram equalization. *Comput Math Methods Med*. 2013;2013:760903.

103. Chen CL, Ishikawa H, Wollstein G, Ling Y, Bilonick RA, Kagemann L, Sigal IA, Schuman JS. Individual A-scan signal normalization between two spectral domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2013;54(5):3463-3471.

104. Zhang Y. Improving the accuracy of direct histogram specification. *Electronics Letters*. 1992;28(3):213-214.

105.   Coltuc D, Bolon P, Chassery JM. Exact histogram specification. *Image Processing, IEEE Transactions on*. 2006;15(5):1143-1152.

106.   Bevilacqua A, Azzari P. A high performance exact histogram specification algorithm. *Image Analysis and Processing, 2007 ICIAP 2007 14th International Conference on*; 2007: IEEE.

107.   Girard MJA, Strouthidis NG, Ethier CR, Mari JM. Shadow Removal and Contrast Enhancement in Optical Coherence Tomography Images of the Human Optic Nerve Head. *Invest Ophthalmol Vis Sci*. 2011;52(10):7738-7748.

108.   Tappeiner C, Barthelmes D, Abegg MH, Wolf S, Fleischhauer JC. Impact of Optic Media Opacities and Image Compression on Quantitative Analysis of Optical Coherence Tomography. *Invest Ophthalmol Vis Sci*. 2008;49(4):1609-1614.

109.   Vermeer KA, van der Schoot J, Lemij HG, de Boer JF. RPE-Normalized RNFL Attenuation Coefficient Maps Derived from Volumetric OCT Imaging for Glaucoma Assessment. *Invest Ophthalmol Vis Sci*. 2012;53(10):6102-6108.

110.   Ishikawa H, Chen CL, Wollstein G, Grimm JL, Ling Y, Bilonick RA, Sigal IA, Kagemann L, Schuman JS. High dynamic range imaging concept-based signal enhancement method reduced the optical coherence tomography measurement variability. *Invest Ophthalmol Vis Sci*. 2013;54(1):836-841.

111.   González RC, Woods RE. Digital image processing: Prentice Hall; 2002.

112.   R Development Core Team. R: A language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing*. 2012.

113.   Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J. OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika*. 2011;76(2):306-317.

114.   Bilonick R. merror: Accuracy and Precision of Measurements. R package version 1.0; 2003.

115.   Heussen FM, Ouyang Y, McDonnell EC, Narala R, Ruiz-Garcia H, Walsh AC, Sadda SR. Comparison of manually corrected retinal thickness measurements from multiple spectral-domain optical coherence tomography instruments. *British Journal of Ophthalmology*. 2012;96(3):380-385.

116.   Chen CL, Ishikawa H, Ling Y, Wollstein G, Bilonick RA, Xu J, Fujimoto JG, Sigal IA, Kagemann L, Schuman JS. Signal normalization reduces systematic measurement differences between spectral-domain optical coherence tomography devices. *Invest Ophthalmol Vis Sci*. 2013;54(12):7317-7322.

117.   Leung CK, Cheung CY, Weinreb RN, Qiu Q, Liu S, Li H, Xu G, Fan N, Huang L, Pang CP, Lam DS. Retinal nerve fiber layer imaging with spectral-domain optical coherence

tomography: a variability and diagnostic performance study. *Ophthalmology*. 2009;116(7):1257-1263, 1263.e1251-1252.

118. 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(4):266-273.

119. Baxi J, Calhoun W, Sepah YJ, Hammer DX, Ilev I, Pfefer TJ, Nguyen QD, Agrawal A. Retina-simulating phantom for optical coherence tomography. *J Biomed Opt*. 2014;19(2):21106.

120. Agrawal A, Connors M, Beylin A, Liang CP, Barton D, Chen Y, Drezek RA, Pfefer TJ. Characterizing the point spread function of retinal OCT devices with a model eye-based phantom. *Biomed Opt Express*. 2012;3(5):1116-1126.

121. Slaney M, Covell M, Lassiter B. Automatic audio morphing. *Acoustics, Speech, and Signal Processing, 1996 ICASSP-96 Conference Proceedings, 1996 IEEE International Conference on*; 1996: IEEE.

122. Sethares WA, Milne AJ, Tiedje S, Prechtl A, Plamondon J. Spectral tools for Dynamic Tonality and audio morphing. *Computer Music Journal*. 2009;33(2):71-84.

123. Ye H, Young S. High quality voice morphing. *Acoustics, Speech, and Signal Processing, 2004 Proceedings(ICASSP'04) IEEE International Conference on*; 2004: IEEE.