Article

Individual Drusen Segmentation and Repeatability and Reproducibility of Their Automated Quantification in Optical Coherence Tomography Images

Luis de Sisternes[1,2,*], Gowtham Jonna[3], Margaret A. Greven[4], Qiang Chen[5], Theodore Leng[4], and Daniel L. Rubin[1,2]

[1] Department of Radiology, Stanford University, Stanford, CA, USA
[2] Department Medicine (Biomedical Informatics), Stanford University, Stanford, CA, USA
[3] Department of Ophthalmology and Visual Sciences, Albert Einstein College of Medicine, Bronx, NY, USA
[4] Department of Ophthalmology, Byers Eye Institute at Stanford, Stanford University School of Medicine, Palo Alto, CA, USA
[5] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
* Currently at Carl Zeiss Meditec, Inc., Dublin, CA, USA

Correspondence: Theodore Leng, MD, MS, Byers Eye Institute at Stanford, 2452 Watson Court, Palo Alto, CA 94303. e-mail: tedleng@stanford.edu
Daniel L. Rubin, MD, MS, Richard M. Lucas Center, 1201 Welch Road, Stanford, CA 94305. e-mail: dlrubin@stanford.edu

Purpose: To introduce a novel method to segment individual drusen in spectral-domain optical coherence tomography (SD-OCT), and evaluate its accuracy, and repeatability/reproducibility of drusen quantifications extracted from the segmentation results.

Methods: Our method uses a smooth interpolation of the retinal pigment epithelium (RPE) outer boundary, fitted to candidate locations in proximity to Bruch's Membrane, to identify regions of substantial lifting in the inner-RPE or inner-segment boundaries, and then separates and evaluates individual druse independently. The study included 192 eyes from 129 patients. Accuracy of drusen segmentations was evaluated measuring the overlap ratio (OR) with manual markings, also comparing the results to a previously proposed method. Repeatability and reproducibility across scanning protocols of automated drusen quantifications were investigated in repeated SD-OCT volume pairs and compared with those measured by a commercial tool (Cirrus HD-OCT).

Results: Our segmentation method produced higher accuracy than a previously proposed method, showing similar differences to manual markings (0.72 ± 0.09 OR) as the measured intra- and interreader variability (0.78 ± 0.09 and 0.77 ± 0.09, respectively). The automated quantifications displayed high repeatability and reproducibility, showing a more stable behavior across scanning protocols in drusen area and volume measurements than the commercial software. Measurements of drusen slope and mean intensity showed significant differences across protocols.

Conclusion: Automated drusen outlines produced by our method show promising accurate results that seem relatively stable in repeated scans using the same or different scanning protocols.

Translational Relevance: The proposed method represents a viable tool to measure and track drusen measurements in early or intermediate age-related macular degeneration patients.

Introduction

Age-related macular degeneration (AMD) is the most common cause of severe vision impairment among patients over 50-years old in the developed world.[1] AMD appears most commonly in early and intermediate nonexudative stages, often not accompanied with noticeable vision loss, and always preceding advanced nonexudative or exudative forms, when visual changes are noticeable and often irreversible. The progression from early or intermediate AMD to an advanced form of the disease (AMD progression) is critical to identify as early as possible. Many patients are unaware of subtle visual changes that are indicators of AMD progression, and prompt
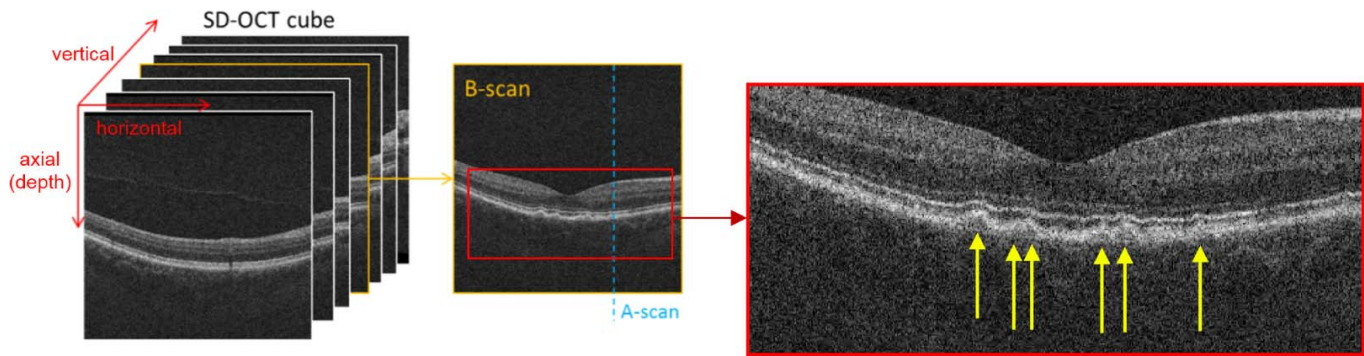
Figure 1. *Left*: SD-OCT cube, with indications of the axial, horizontal, and vertical directions, and B-scan and A-scan nomenclature. *Right*: Detail B-scan where location of drusen is indicated with *yellow arrows*.

intervention with pharmacologic treatments can greatly improve visual outcomes.[2] Early and intermediate AMD are normally characterized by the presence of extracellular material that accumulates between Bruch's membrane and the retinal pigment epithelium (RPE) known as drusen, whose presence and properties are the most widely used risk discriminator determining future AMD progression,[3–6] among recent discoveries in genetic,[7] behavioral, environmental, and other phenotypic characterizations. Drusen can present numerous distinctive properties,[8] and their quantitative characteristics and evolution over time can potentially be a powerful biomarker and indicator of AMD progression that should be investigated in depth.[9]

There is positive correlation in the number, size, and extent of drusen seen in color fundus photographs (CFP) with risk of AMD progression.[3–6] However, drusen can sometimes be difficult to identify in the photographs, and their quantification by visual inspection is prone to errors and reader variability. Spectral-domain optical coherence tomography (SD-OCT),[10] an in-vivo imaging method capable of resolving cross-sectional retinal substructures, has potential value in better characterizing RPE deformities. SD-OCT can visualize and quantify the changes seen at different stages of AMD,[11] because the RPE, the site of many of those pathological changes, is well visualized using this imaging technique. Drusen normally appear in SD-OCT images as "bumps" in an otherwise smoothly curved RPE layer (as seen on Figure 1, with drusen locations indicated by yellow arrows), enabling their identification and quantification. Among other research groups, our group has recently proposed the observation of a number of drusen SD-OCT quantitative imaging features as predictors for AMD progression.[9]

In order to quantitatively characterize drusen via

SD-OCT, their presence, location, and extent within the images must first be determined. A single SD-OCT volume typically contains from 100 to 200 two-dimensional (2D) high-resolution images (B-scans), which makes manual annotation of drusen boundaries a very time-consuming task, with considerable inter- and intrareader differences.[12–14] Consequently, most research in the quantification of SD-OCT features involves an automated or semiautomated segmentation approach, where the process can be less time-consuming, more reproducible, and parallelized. Many automated or semiautomated retinal layer segmentation methods[15–24] and a few methods aimed to the segmentation of drusen regions[25–30] have been proposed previously. Recently, commercial SD-OCT instruments have also incorporated automated software methods to characterize drusen volumes and areas.[31] However, drusen descriptors still rely in the characterization of RPE lifting properties across the entire macula region rather than the detection and identification of individual drusen (or clusters of them), not permitting the extraction of a number of features from individual drusen that may be significant in AMD progression, such as the number of drusen, individual drusen volume, and area, distribution, or reflectivity properties.[9] Apart from accuracy requirements, it is also important to assess the repeatability and reproducibility of the quantitative features that may be extracted to characterize drusen from the resulting outlines.

Our group recently proposed a rather simplistic but accurate algorithm for the segmentation of drusen regions based in a 2D identification of the RPE layer.[14] In this paper, we expand and improve our work in drusen segmentation by employing the results from a three-dimensional (3D) retinal layer segmentation algorithm[16] and the identification of individual drusen. We evaluate the accuracy of the method by
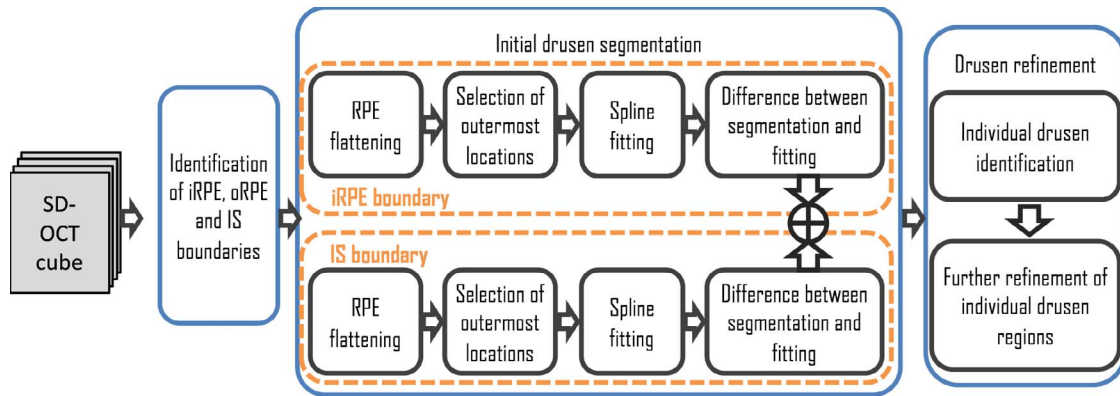
Figure 2. Diagram of the drusen segmentation method.

comparison to manual markings. We also asses repeatability and reproducibility of a set of quantitative drusen features related to AMD progression[9] that are extracted by using our proposed method and drusen area and volume that are automatically generated by a commercial system, comparing them between repeated measurements obtained during a single visit in several patients.

Methods

Automated Drusen Segmentation

Our drusen segmentation algorithm follows a similar idea introduced by our group in a previous publication,[14] but expanding and improving the method using recent work in retinal layer segmentation and a fully 3D approach. The algorithm main steps are illustrated in Figure 2, which include the segmentation of the RPE inner and outer boundaries (iRPE and oRPE, repectively) and the boundary delimiting the inner segment location (IS), using an automated segmentation method.[16] An initial segmentation of drusen is generated based in the main idea of fitting the segmented iRPE and IS boundaries to a surface that would preserve the typical curvature of the healthy RPE complex while eliminating the curvatures produced by drusen. Because drusen can be typically observed as small "bumps" in the RPE complex normal curvature, finding the regions between the segmented boundaries and their fitted versions produces the initial drusen segmentation. We analyze the curvature in both iRPE and IS boundaries to improve the sensitivity in drusen detection and minimize the influence of possible errors in boundary segmentation, as the iRPE boundary may be difficult to identify in some SD-OCT scans where the cone inner and outer segments (COST) and RPE cannot be discriminated accurately. A series of refinement steps follow to identify the extent of each individual druse and eliminate regions falsely detected as drusen. These main steps are further described in the following subsections.

Identification of iRPE, oRPE, and IS Boundaries

We employed our recently proposed 3D retinal layer segmentation algorithm to automatically outline the iRPE, oRPE, and IS boundaries.[16] The algorithm consists of a two-step process: (1) initial estimation of the retinal layers axial location based on intensity and gradient statistics of the images, and (2) refinement of this estimation though an iterative process. The iterative refinement employs a weighted median (WM) filter, which was formulated and adapted to the particular case of retinal layer segmentation in SD-OCT cubes, together with the consideration of anatomical constraints common to every eye and their expected smooth behavior. The resulting outlines obtained from this method in the images employed here were reviewed to verify their correctness. An example of the segmentation results of such boundaries is shown in Figure 3A.

Initial Drusen Segmentation

We followed the same independent process for the initial segmentation of drusen using the resulting iRPE and IS boundaries (Fig. 2). The main idea behind drusen identification is fitting a surface that follows each boundary outermost locations (with higher axial value) and smoothes out the presence of the "bumps" that indicate drusen. Drusen can then be initially identified as the regions located between the segmented boundary and this fitted surface. The steps for obtaining a fitted surface of these characteristics are indicated in Figure 2 and examples of resulting
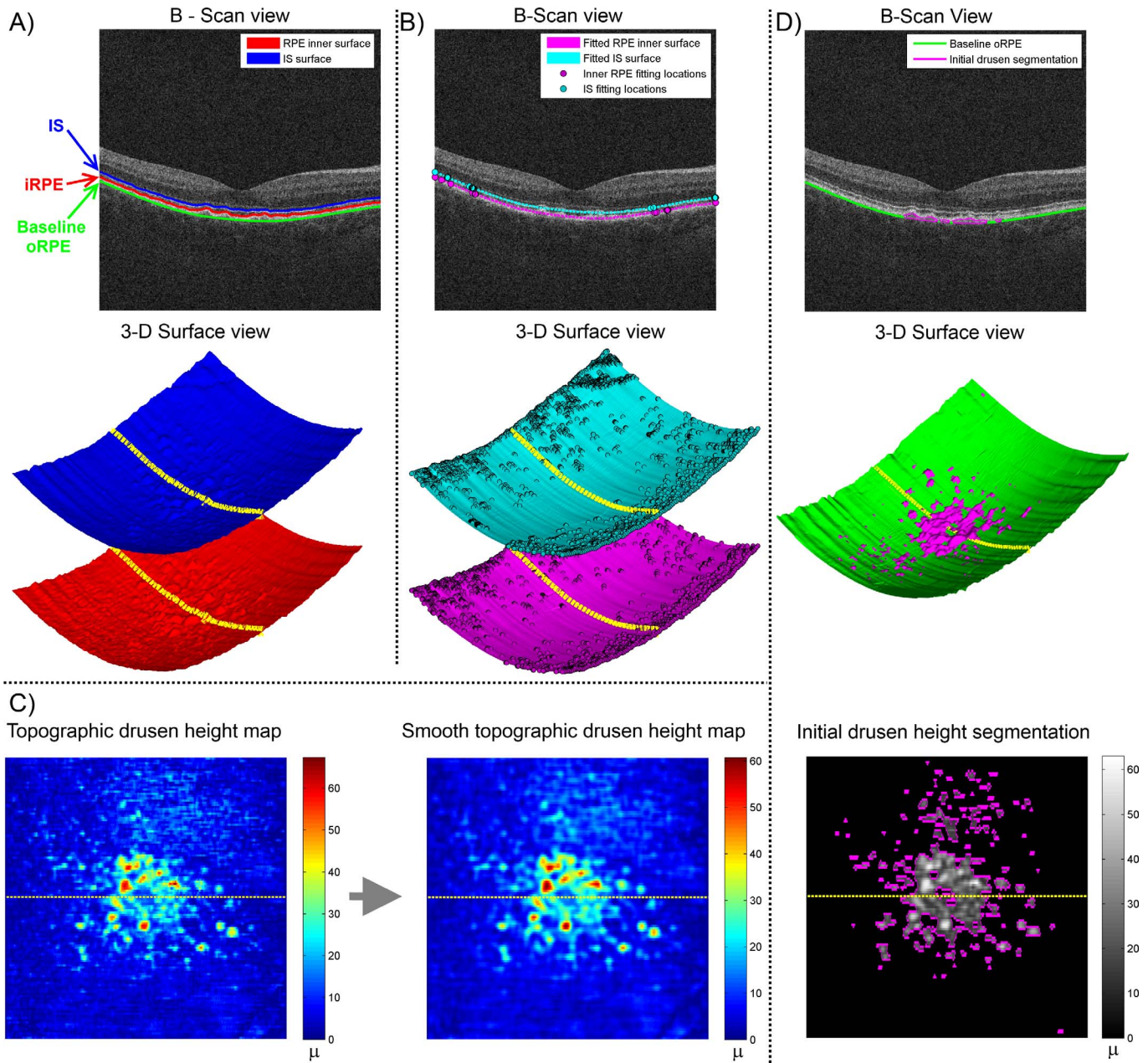
Figure 3. Steps in initial drusen segmentation. (A) Segmented surfaces in example B-scan (top) and 3D surface view (*bottom*, where axial distance between boundaries is exaggerated for displaying purposes). (B) Result from candidate location selection and fitting of the iRPE and IS boundaries in example B-scan (*top*) and 3D surface view (*bottom*). (C) Generated topographic drusen height map and the result after filtering. (D) Results from initial drusen segmentation (in magenta) in example B-scan (*top*), in 3D surface view (*middle*) and in topographic view (*bottom*). *Yellow markings* in all images correspond to the location of the B-scans shown on top.

surfaces are illustrated in Figure 3B. Our method first selects a set of candidate locations from each boundary, selected as those that are most likely not affected by drusen and following a normal healthy RPE curvature, and then fits a thin-plate spline surface[32] (tps) to these candidate locations.

Candidate locations are those following a convex shape with respect to the normal RPE curvature expected in healthy eyes. In order to account for this expected normal curvature, candidate selection was preceded by a "flattening" step, using a smooth version of the oRPE boundary as offset (generated by using a diffusion regularization fitting operation, with fitting tools available online for download).[33] That is,

the axial location of the both iRPE and IS boundaries are remapped according to its location respect to a "baseline RPE" location, which is common to both surfaces (RPE flattening step in Fig. 2). The candidate locations are then selected independently for the iRPE and IS boundaries as those following a convex shape in each B-scan, which selection can be done with the following process:

Step 1: A location set is created by ordering the "flattened" boundary locations in decreasing depth.

Step 2: Candidate locations are selected one by one in order from the location set. As a new candidate is selected, all remaining locations with horizontal position falling between the intervals formed by candidates already selected are deleted from the location set.

Step 3: Repeat Step 2 until the location set is empty.

The selected candidate locations are remapped to their original axial position by inverting the "flattening" step. Two different tps surfaces are then fitted to the candidate locations, for those selected from the iRPE and IS boundaries, respectively (example shown in Fig. 3B). A topographic map of drusen height is then formed by the difference between the segmented iRPE and IS boundaries and their corresponding fitted surfaces, selecting the maximum between the two. As drusen height is expected to present a smooth undulatory behavior throughout the span of the SD-OCT scan, the topographic drusen height map is smoothed using a 2D fourth degree polynomial Savitzky-Golay filter of 0.25-mm window length.[34] This choice of filtering reduces possible noise expected in the segmentation results by limiting the appearance to only one drusen in a span of 0.0156 mm^2. An example of the topographic drusen height maps before and after this filtering operation is shown in Figure 3C.

The initial drusen segmentation is then formed by considering drusen height in the resulting topographic map with respect the baseline RPE. Regions presenting an axial height of less than 15 μm are eliminated from the segmentation. The remaining regions having a maximum drusen height under 20 μm are also eliminated. The reasoning behind this threshold value is that we are expected to find a large number of low separation locations due to fitting differences (noise) and not caused by the presence of drusen. Considering the typical thickness of the RPE layer and the resolution of the SD-OCT system, we observed than while limiting our methods to detect drusen locations

of at least 15 μm of height, the results using this constraints were satisfactory. An example of the resulting initial drusen segmentation can be observed in Figure 3D.

Individual Druse Identification and Refinement

A series of refinement steps follow to separate individual drusen and further eliminate possible falsely detected regions in the segmentation results. These steps are illustrated in the example shown in Figure 4 and are as follows:

1. Individual drusen regions are first identified using the Watershed transform[35] of the topographic drusen height map. This process separates the initial regions (shown in Fig. 4A) into those characterized by pronounced local increases of height (separation shown in Fig. 4B), a characteristic to expect in individual drusen.

2. The topographic locations in the ridges or borders resulting after the watershed operation are added to the nearest identified individual region. Those resulting individual regions with a maximum height under 20 μm (minimum druse peak height considered in this work), are then further analyzed pixel by pixel, with each pixel either being eliminated from the segmentation or added to a neighboring region that follows this height requisite: a pixel gets added to the neighboring region of weighted centroid (weighted by the height map) at minimum distance, as long as the pixel is within the neighboring region's maximum extent (major axis). If no neighboring regions of such characteristics can be found, the pixel location is removed from the segmentation. The reasoning behind this process is to further analyze individual regions with a height peak lower than the given threshold (20 μm), adding part of their pixels to neighboring regions and removing those that are not within the extent of neighboring regions. The result of this process can be seen in Figure 4C.

3. The resulting individual regions are further processed by a morphological opening operation with a disk of diameter equal to the region's minor axis in order to smooth the edges of each individual druse in a topographic view. Results after this process can be observed in Figure 4D.

An example of the final results produced after this individual drusen characterization and refinement is shown in Figure 4E. Comparing Figure 4A and
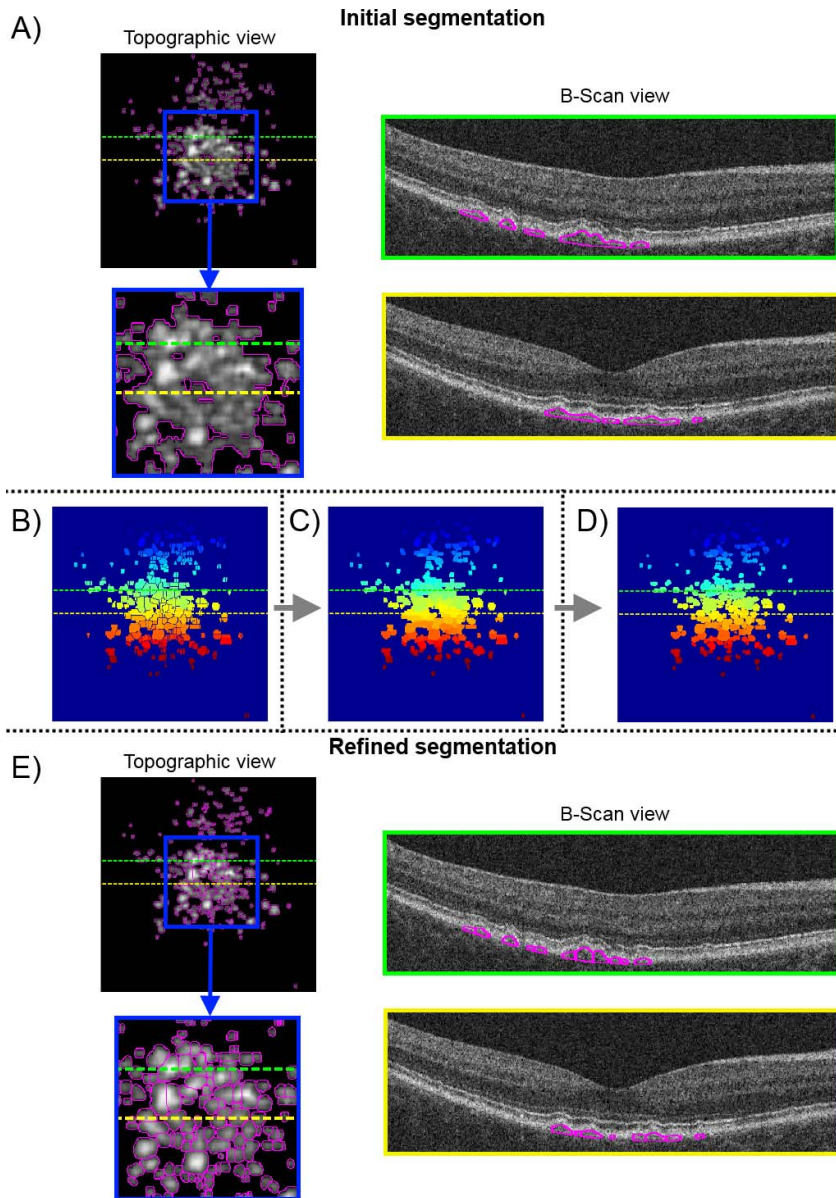
Figure 4. Individual druse identification and refinement in example scan. The *green* and *yellow lines* shown in the topographic views indicate the location of the B-scans, outlined in *green* and *yellow*, respectively. (A) Initial drusen segmentation (*magenta outline*) in topographic height view (the *blue square* indicates the detail location shown below) and in two example B-scans. (B–D) Topographic height view after the watershed transform, further processing of individual regions of low height, and morphological opening of individual regions, respectively. The different colors indicate different individual drusen regions. (E) Refined drusen segmentation (*magenta outline*) in topographic height view (the *blue square* indicates the detail location shown below) and in two example B-scans.

Figure 4E we can observe how the initial regions, containing a conglomerate of drusen, are separated into individual regions outlining each of the individually detected drusen.

Drusen Quantification

In previous work we developed methods to extract a series of quantitative features from the drusen segmentations in SD-OCT scans that have shown utility as predictors of AMD progression.[9] These features include characteristics of drusen shape and geometry, total drusen area, 3D volume of drusen, number of drusen in cube, mean area and volume per detected drusen, maximum drusen height, extent of retinal area affected by drusen, drusen density in affected area, drusen slope, drusen reflectivity, and texture properties in the region delimited by drusen

(mean and standard deviation [SD] of reflectivity inside drusen). This set of features was extracted from the segmentations obtained by our automated method to study their repeatability and reproducibility. The complete list of extracted features characterizing is enumerated in Table 2. For a detailed description of each feature we refer to the previous work.[9]

Image Datasets and Collected Analysis

We obtained three datasets to evaluate our automated drusen segmentations and quantifications. The first two datasets (Datasets A and B) were used to evaluate the accuracy of the drusen segmentations; they are the same as those used in previous drusen segmentation work.[14] The third dataset (Dataset C), was used to study the repeatability and reproducibility of drusen quantifications derived from our segmentations, and it contains repeated scans from a separate set of eyes. All eyes included in the three datasets presented drusen of several characteristics (including both sub-RPE cuticular drusen and sub-retinal deposits above the RPE, also known as pseudodrusen) and were diagnosed with early or intermediate nonneovascular AMD. Eyes presenting pigment epithelial detachments (PED) were excluded from the study. The study was conducted in accordance with the ethical standards stated in the 1964 Declaration of Helsinki. All subjects underwent complete eye examinations, including visual acuity tests, CFPs, and SD-OCT imaging, and nonneovascular AMD diagnosis was established by an experienced ophthalmologist. All eyes were designated as having either dry AMD belonging to the Age-Related Eye Disease Study (AREDS) categories 2 or 3,[4] presenting a combination of multiple small drusen and few intermediate drusen or a combination of extensive intermediate drusen and at least one large drusen. Eyes with significant media opacity resulting in poor OCT image quality or evidence of geographic atrophy or chorodial neovascularization were excluded from the study.

All SD-OCT scans included in the study were collected using the same instrument model (Cirrus HD-OCT; Carl Zeiss Meditec, Inc., Dublin, CA). Scans were acquired using the Cirrus system 512×128 macular cube protocol or the 200×200 macular cube protocol, producing $1024 \times 512 \times 128$ volumetric data (128 B-scans with 512 A-scan per B-scan and 1024 equally spaced axial lines per A-scan; SD-OCT nomenclature is displayed in the example shown in Fig. 1) or $1024 \times 200 \times 200$ volumetric data (200 B-scans with 200 A-scan per B-scan and 1024 equally spaced axial lines per A-scan),

respectively. Both protocols corresponded to a $2 \times 6 \times 6$ mm. region around the fovea center. The analysis of each dataset varied as follows:

Dataset A: Included four SD-OCT volumetric data-sets from four different eyes in three patients acquired using the 512×128 scan protocol. Drusen in each B-scan in the dataset (a total of 512 B-scans) was manually segmented by two different experts to analyze interobserver agreement. To assess intraobserver variation, each expert also segmented each image twice in two different sessions at least 2 months apart. Only four cubes were considered in this dataset because of the enormity of the task of creating manual markings. Each expert created drusen outlines in each of the 128 B-scan images in each cube. Automated segmentations produced by the method presented here and from our previous method[14] were generated for comparison.

Dataset B: Included 143 SD-OCT volumetric image datasets from 143 eyes in 99 patients acquired using the 512×128 scan protocol. Manual drusen segmentations were drawn by a third expert on one B-scan in each of the 143 eyes, selecting the most representative B-scan for each cube as the one that contained the largest drusen. While the automated method produces 3D drusen segmentation throughout the SD-OCT cube, only one B-scan for each cube was manually segmented and considered in this dataset, with the goal of evaluating the method over a large number of different eyes, while keeping the time intensive task of drawing drusen outlines within practical limits. Automated segmentations produced by the method presented here and by our previous method[14] were generated for comparison.

Dataset C: Included 45 repeated SD-OCT volumetric dataset pairs from 45 eyes in 27 patients using a 512×128 scan protocol. For 23 of these scan pairs, we also collected repeated scan pairs during the same clinic visit using the 200×200 scan protocol. Repeated SD-OCT data from the same subjects were collected within 5-

minute intervals in the same visit using identical instrument and settings. The set of drusen quantitative features were extracted from the segmentation results generated by our method in all the datasets (136 in total). We also collected drusen area and volume both in a 3-mm radius circle and in a 5-mm radius circle from the center of the fovea (four features) as measured by the commercial software in the acquisition system (Cirrus HD-OCT Review Software, version 7.0.1.290) for all the SD-OCT data. These same four features were also computed as produced by our segmentation method for comparison. Repeatability of the extracted features was assessed by comparison of the values between the repeated scans of the same eyes using the same scanning pattern. Reproducibility was computed by comparing the feature values as obtained from the same eyes using different scanning protocols.

Evaluation of Accuracy

We considered the absolute drusen area difference (ADAD) and the OR as metrics to quantify the segmentation differences between two segmentation results (either by two different readers, at different sessions or automatic versus manual segmentation). ADAD quantifies the absolute differences in drusen area measured at each segmented B-scan, and was measured both in microns and in percentages relative to the mean area of the two outlines compared. OR, computed on each B-scan, was defined in the same way as the Jaccard index, and it quantifies the ratio of area between the intersecting region of two outlines and their union, taking values between 0 (indicating no pixels in common between two outlines) and 1 (indicating that two outlines are exactly the same). The correlation of drusen area measurements in each B-scan produced by two segmentation results was also analyzed by computing Pearson's correlation coefficients,[36] and the significance of their differences was tested using the paired Mann-Whitney U test, which is the nonparametric counterpart of the paired sample t-test.[37] The justification of using this nonparametric test over a paired sample t-test was that we could not assume the data (drusen areas segmented in B-scans) follow a normal distribution, and we might expect large tails in their distribution.

A diagram summarizing the evaluations undertaken is shown in Figure 5A. Intra- and interreader variability were first analyzed by comparing the above metrics from the outlines generated by the same reader at two different sessions and from the two readers, respectively, in Dataset A. Intrareader variability was analyzed for both manual readers independently. When comparing outlines from the two different readers (inter-reader), we considered both segmentations produced at the two sessions for each reader. Comparison between manual markings and the automated segmentations from our method and the previous approach[14] was made using both Dataset A and Dataset B independently with the metrics above. For Dataset A, automated results were compared with the average segmentation between readers and sessions, averaging their outline placement in each A-scan.

Evaluation of Repeatability and Reproducibility

We assessed the repeatability and reproducibility of the drusen features extracted from our automated segmentation method and those collected from the Cirrus software by computing a coefficient of repeatability (CR) from the results produced in the repeated SD-OCT scans in Dataset C. CR was measured as the percentage difference observed in a feature value between two repeated scans with respect to its average in the repeated scans. Statistical significance of the differences in the extracted features between repeated scans was also assessed using the U test. Resulting P values were corrected using a multiple hypothesis testing by estimating the false-positive discovery rate q-values.[38] Differences having a q-value less than 0.05 were considered statistically significant. A diagram for the comparisons evaluated is shown in Figure 5B.

Repeatability for the 512×128 and 200×200 scanning protocols was evaluated by analyzing differences in repeated scans collected using the same protocol. Reproducibility across scanning protocols was evaluated by analyzing the differences in repeated scans obtained using the two different protocols, employing the average feature value obtained from the repeated SD-OCT scan pairs at each protocol. The set of eyes where reproducibility was analyzed included those eyes with scans using both protocols, which corresponded to the same set of eyes where repeatability for the 200×200 protocol was analyzed.

A) Evaluation of Accuracy

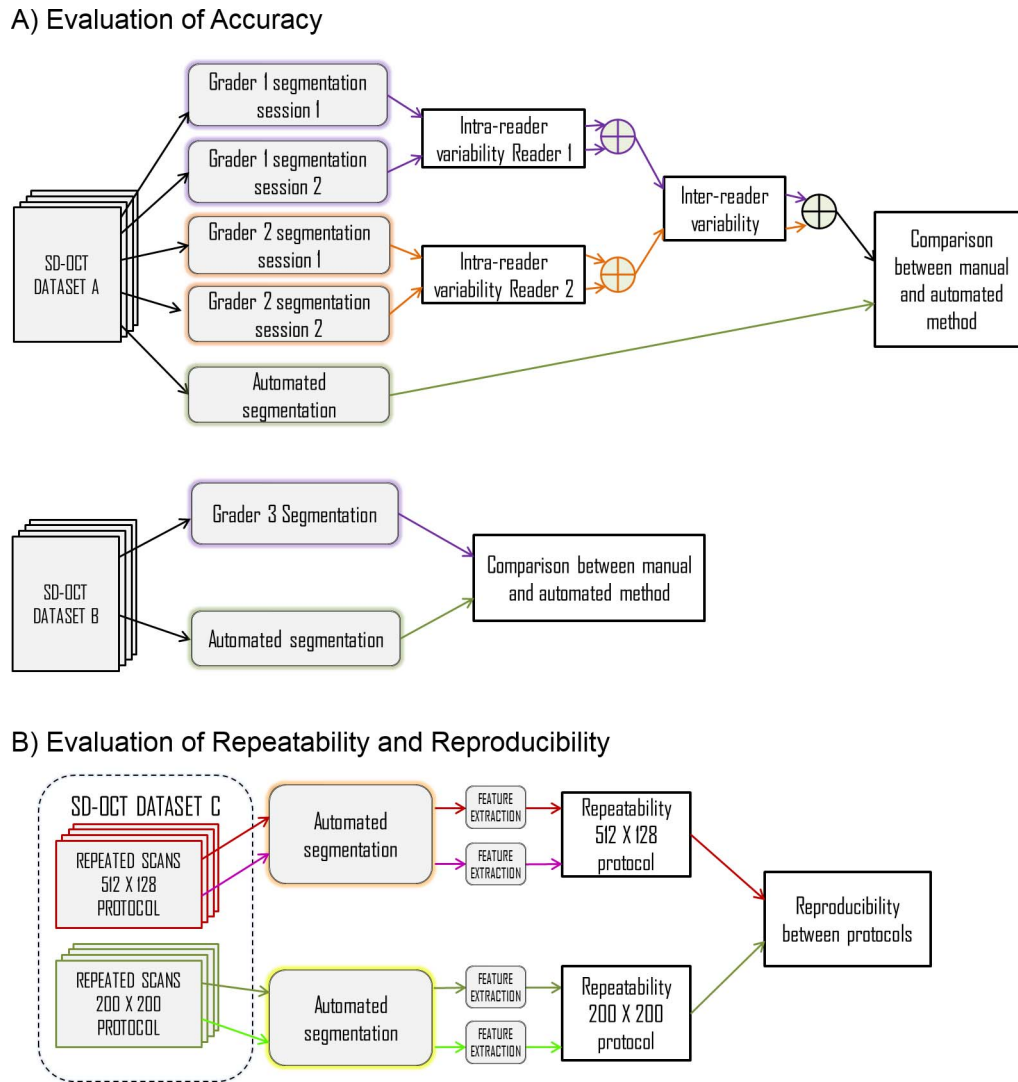## B) Evaluation of Repeatability and Reproducibility

Figure 5. (A) Diagrams of comparisons between manual and automated segmentation to study the accuracy of our methods. (B) Diagram of comparison between repeated scans to study the repeatability and reproducibility of drusen features extracted from the automated drusen segmentation.

Results

Evaluation of Accuracy

Table 1 shows the agreement in inter-reader and intra-reader segmented drusen area in terms of correlation coefficients, U test P values, ADAD, and OR for Dataset A. The interreader differences were slightly higher for the first expert, but reader agreement was relatively high overall, showing very high correlation values (0.96–0.98) and consistent OR (0.77–0.79). No statistically significant differences were observed when comparing drusen area measurements within or between readers. Table 1 also summarizes the agreement between the average reader segmentations and both automated segmentations produced by an earlier method (Aut. Chen)[14] and the method presented here (Aut. Seg., in boldface) in Dataset A and Dataset B. For the smaller dataset (Dataset A), the correlation coefficient between automated methods and manual reader segmentations was very high for both methods, being higher for our improved method (0.99) and similar to those observed within-experts and between-experts. ADAD and overlap ratio values between automated and manual segmentations were also very similar to those within and between readers, with differences within the measured SDs. Average overlap ratios in Dataset A showed a slightly lower value for our method than

Table 1. Evaluation of Drusen Segmentation Accuracy

Methods Compared	# Eyes / B-Scans Evaluated	CC	*U*-test, *P* Value	ADAD, μm Mean (SD)	ADAD, %, Mean (SD)	OR, Mean (SD)
Dataset A						
Intrareader 1	4/512	0.97	0.845	9.25 (5.93)	18.42 (14.1)	0.77 (0.09)
Intrareader 2	4/512	0.98	0.062	6.14 (6.03)	11.36 (12.4)	0.79 (0.09)
Interreader	8/1024	0.96	0.083	8.11 (7.52)	14.94 (13.45)	0.77 (0.09)
Aut. Chen - Avg. R.	4/512	0.96	0.776	7.89 (6.46)	16.05 (13.8)	0.74 (0.09)
Aut. Seg. – Avg. R.	4/512	**0.99**	0.108	**6.16 (5.74)**	**11 (9.51)**	**0.72 (0.09)**
Dataset B						
Aut. Chen - Avg. R.	143/143	0.92	< 0.01	23.38 (12.45)	37.78 (21.39)	0.62 (0.11)
Aut. Seg. – Avg. R.	143/143	**0.95**	0.112	**13.82 (13.06)**	**19.13 (14.94)**	**0.68 (0.1)**

Aut. Chen: Indicates drusen outlines generated using a previous method.[14] Avg. R: Indicates drusen outlines generated by averaging the manual readers and sessions. Aut. Seg: Indicates drusen outlines generated using our improved segmentation method. cc: Pearson's correlation coefficients.

for the previous method (0.72 vs. 0.74), although this difference was very small when compared with the observed SD (0.09 in both cases). For the larger dataset (Dataset B), the differences found between automated and manual segmentations were in the same ranges as observed in Dataset A for our improved segmentation method and quite higher for the previous method (ADD% of 19.13 ± 14.94 vs. 37.78 ± 21.39). While the *P* values indicate that statistical differences could not be claimed for the previous method in the first dataset, they were significant in the second dataset. On the other hand, such differences were not significant for either dataset when evaluating our improved segmentation method. These results suggest that the discrepancies observed between the hand-drawn segmentations are comparable to those observed between the automated method presented here and manual readers. Our method also shows improvements over the previous method in the larger dataset.

Evaluation of Repeatability and Reproducibility

Table 2 summarizes the repeatability and reproducibility evaluation of drusen features extracted using the drusen segmentations obtained by our method (additional Bland-Altman plots for each of the features and comparisons are shown in Figs. 7 and 8 of the Appendix). All features showed relatively low average percentage differences (CR) between repeated scans using the same scanning protocol (in the range of 2.54%–13.2% and 2.15%–16.89% for the 512 × 128 and 200 × 200 protocols) and when comparing between protocols (1.56%–13.81%). Although still

showing low CR values, the only statistical differences were observed between mean drusen slope and mean reflectivity inside drusen regions (which measures the mean normalized intensity recorded in SD-OCT scans) across different scanning protocols. These results seem to be due to the differences observed when evaluating drusen shape at different protocols (a larger number of B-scans should allow a more detailed characterization in the vertical axis) and the intrinsic differences observed in intensity values between the two different protocols.

Table 3 summarizes a comparison of repeatability and reproducibility of drusen area and volume computed using the drusen segmentations obtained by our method (Aut. Features) and from commercial software (Cirrus features) (additional Bland-Altman plots for each of the features and comparisons are shown in Fig. 9 of the Appendix). Overall, the quantifications using our method were more repeatable than the ones generated by the commercial software when the 512 × 128 scanning protocol was used. It is important to note that the commercial software produced maximum and minimum CR of 200 and 0, respectively, when using this protocol. This seems to be an effect of a limitation in the commercial software when identifying drusen regions at lower vertical resolutions (128 B-scans in this case), estimating a drusen area and volume of 0 for some scans that contained small drusen. An example of this effect, shown in Figure 6, displaying the repeated scans with highest differences as measured by our method. The ability of our method to recognize small drusen can be observed in the higher average area and volume measurements.

Table 2. Repeatability and Reproducibility of Quantitative Drusen Features Generated by Our Improved Segmentation Method
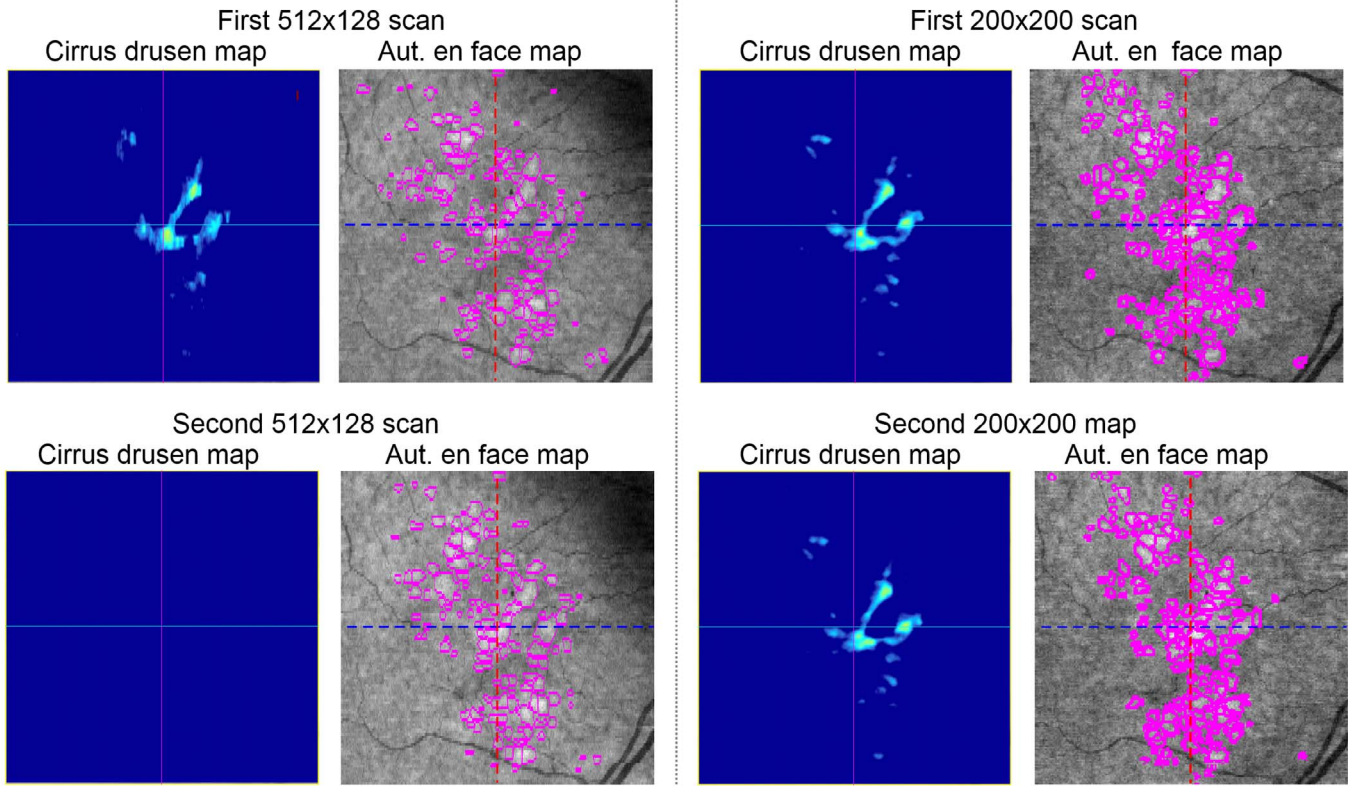
	Repeatability 512 × 128 (N = 45)			
	Feature Values Mean (SD)	CR %.		
Drusen Feature Analyzed		Mean (SD)	Range	q-Value
Number of drusen in cube	103.79 (57.11)	11.13 (10.62)	0–48.33	0.99
Mean volume per drusen (mm³/drusen)	0.004 (0.001)	8.16 (5.75)	0.26–21.16	0.99
Drusen aggregate volume (mm³)	0.373 (0.214)	9.43 (10.93)	0.66–43.26	0.99
Mean area per drusen (mm²)	0.049 (0.013)	7.38 (4.89)	0.09–17.56	0.99
Drusen aggregate area (mm²)	5.038 (2.693)	9.41 (11.53)	0.01–47.1	0.99
Drusen density in projection image	0.301 (0.126)	15.75 (16.72)	0.29–79.75	0.99
Extent of drusen affected area (mm²)	16.77 (6.11)	13.2 (15.72)	0.29–66.4	0.99
Maximum drusen height (mm)	0.145 (0.031)	6.64 (7.8)	0–35.62	0.99
Mean drusen slope	0.581 (0.065)	4.04 (3.13)	0.03–12.75	0.99
Mean reflectivity inside drusen region	0.383 (0.035)	5.56 (4.95)	0.42–23.67	0.99
SD reflectivity inside drusen region	0.114 (0.006)	2.54 (2.51)	0.02–9.73	0.99

Table 2. Continued

	Repeatability 200 × 200 (N = 23)			
	Feature Values Mean (SD)	CR % mean (SD) Maximum, Minimum		
Drusen Feature Analyzed		Mean (SD)	Range	q-value
Number of drusen in cube	131.87 (44.29)	16.89 (15.13)	1.29–50	1
Mean volume per drusen (mm³/drusen)	0.004 (0.001)	11.66 (8.49)	0.48–29.17	1
Drusen aggregate volume (mm³)	0.497 (0.172)	7.07 (8.51)	0.36–33.13	1
Mean area per drusen (mm²)	0.050 (0.007)	10.07 (6.85)	0.31–25.72	1
Drusen aggregate area (mm²)	6.424 (2.073)	8.39 (10.35)	0.02–35.41	1
Drusen density in projection image	0.328 (0.076)	16.26 (11.93)	0.85–42.32	1
Extent of drusen affected area (mm²)	19.708 (4.742)	16.61 (14.14)	0.73–50.19	1
Maximum drusen height (mm)	0.170 (0.029)	3.07 (2.97)	0–11.61	1
Mean drusen slope	0.498 (0.043)	4.73 (2.72)	0.47–9.33	1
Mean reflectivity inside drusen region	0.336 (0.028)	3.74 (2.88)	0.34–9.75	1
SD reflectivity inside drusen region	0.114 (0.007)	2.15 (1.5)	0.32–5.99	1

Table 2. Continued

	Reproducibility (N = 23)		
	CR % mean (SD) Maximum, Minimum		
Drusen Feature Analyzed	Mean (SD)	Range	q-Value
Number of drusen in cube	13.81 (13.2)	0–47.21	0.3
Mean volume per drusen (mm³/drusen)	9.91 (7.71)	0.47–25.88	0.75
Drusen aggregate volume (mm³)	8.82 (5.08)	0.6–18.91	0.55
Mean area per drusen (mm²)	8.18 (5.94)	0.2–23.44	0.20
Drusen aggregate area (mm²)	7.94 (5.88)	0.21–22.7	0.66
Drusen density in projection image	8.78 (6.16)	0.23–26.22	0.20
Extent of drusen affected area (mm²)	13.75 (12.25)	1.83–48.68	0.12
Maximum drusen height (mm)	7.41 (8.23)	0–29.21	0.12
Mean drusen slope	6.98 (3.26)	0.13–13.3	0.002
Mean reflectivity inside drusen region	7.09 (3.29)	1.27–11.81	0.002
SD reflectivity inside drusen region	1.56 (1.26)	0.01–4.46	0.66

Figure 6. Example of repeatability and reproducibility analysis for a single eye where differences in area at 5-mm across protocols using our method were the highest. Results for the repeated scans at 512 × 128 and 200 × 200 protocols are shown. The Cirrus maps display drusen elevation maps as produced by Cirrus software. The Aut. en face maps represent topographic projection maps of the RPE complex, where outlines of the drusen segmentation results produced by our method are superimposed. The table summarizes the area and volume measurements for each scan, repeatability and reproducibility (in CR%) generated by Cirrus and by our method.

Feature	First measurement 512x128	Second measurement 512x128	First measurement 200x200	Second measurement 200x200	Repeatability 512x128	Repeatability 200x200	Reproducibility
Aut. Features							
Area at 3mm (mm²)	2.84	2.89	3.50	2.90	1.74	18.75	11.05
Area at 5mm (mm²)	4.65	4.87	5.68	5.29	4.62	7.11	14.15
Volume at 3mm (mm³)	0.181	0.184	0.229	0.191	1.64	18.10	14.01
Volume at 5mm (mm³)	0.291	0.304	0.359	0.335	4.37	6.92	15.36
Cirrus Features							
Area at 3mm (mm²)	0.9	0	0.8	0.8	200	0	56
Area at 5mm (mm²)	0.9	0	0.9	0.8	200	11.76	65.54
Volume at 3mm (mm³)	0.03	0	0.03	0.02	200	40	50
Volume at 5mm (mm³)	0.03	0	0.03	0.03	200	0	66.67

Measurements had greater repeatability for both methods when employing the 200 × 200 protocol than with the 512 × 128 protocol, especially for the commercial software: our segmentation method had small differences (mean CR between 6.57%–8.06%), and the commercial software presented an excellent repeatability (mean CR between 2.15%–4.15%). When evaluating reproducibility across protocols, our method had greater reproducibility (mean CR 6.21%–9.47%) than the commercial software (14.13%–14.66%).

Discussion

It has been previously shown that cumulative drusen area and volume measured in SD-OCT images may be potential disease biomarkers of AMD status and progression, and reproducibility of such quantifications has been partly studied.[31,39,40] As retinal drusen can also be categorized into different types according to their shape, predominant internal reflectivity, homogeneity, and presence of overlying

Table 3. Repeatability and Reproducibility of Drusen Area and Volume Measurements Generated Using our Improved Segmentation Method (Aut. Features) and That Produced by Commercial Software (Cirrus Features)

| | | Repeatability 512 × 128 (N = 45) | | |
| | Feature Values, | CR % | | |
Drusen Feature Analyzed	Mean (SD)	Mean (SD)	Range	q-Value
Aut. Features				
Area at 3 mm (mm^2)	2.778 (1.2727)	9.61 (13.55)	0.07–70.89	0.99
Area at 5 mm (mm^2)	4.4287 (2.358)	9.75 (12.73)	0–59.23	0.99
Volume at 3 mm (mm^3)	0.212 (0.108)	9.55 (13.01)	0.31–70.99	0.99
Volume at 5 mm (mm^3)	0.329 (0.188)	9.63 (11.93)	0.27–54.31	0.99
Cirrus Features				
Area at 3 mm (mm^2)	1.232 (1.024)	18.9 (44.4)	0–200	0.99
Area at 5 mm (mm^2)	1.831 (1.524)	30.46 (59.85)	0–200	0.99
Volume at 3 mm (mm^3)	0.081 (0.100)	19.62 (47.1)	0–200	0.99
Volume at 5 mm (mm^3)	0.119 (0.169)	22.68 (47.73)	0–200	0.99

Table 3. Continued

| | Repeatability 200 × 200 (N = 23) | | | | Reproducibility (N = 23) | | |
| | Feature Values, | CR % | | | CR % | | |
Drusen Feature Analyzed	Mean (SD)	Mean (SD)	Range	q-Value	Mean (SD)	Range	q-Value
Aut. Features							
Area at 3 mm (mm^2)	3.727 (0.636)	7.2 (7.58)	0.03–34.92	1	6.33 (3.95)	0.16–13.63	0.66
Area at 5 mm (mm^2)	5.893 (1.808)	8.06 (9.68)	0.12–35.32	1	6.21 (4.08)	0.36–14.19	0.75
Volume at 3 mm (mm^3)	0.316 (0.080)	6.57 (7.46)	0.22–34.96	1	9.47 (5.23)	0.81–18.51	0.49
Volume at 5 mm (mm^3)	0.465 (0.156)	6.98 (7.93)	0.04–34.54	1	7.48 (4.3)	0.32–16.44	0.54
Cirrus Features							
Area at 3 mm (mm^2)	1.791 (0.760)	2.15 (3.09)	0–10.53	1	14.13 (41.54)	0–200	0.75
Area at 5 mm (mm^2)	2.474 (1.305)	3.66 (4.29)	0–11.76	1	14.19 (41.68)	0–200	0.75
Volume at 3 mm (mm^3)	0.109 (0.074)	4.15 (8.86)	0–40	1	14.55 (41.44)	0–200	0.75
Volume at 5 mm (mm^3)	0.142 (0.098)	3.33 (4.5)	0–11.76	1	14.66 (41.88)	0–200	0.75

hyperreflective foci,[8] characterization of individual drusen properties is of interest and can increase the prediction accuracy when analyzing AMD progression risk.[9] For example, the imaging phenotype of an eye with a small number of large drusen should be different from that of an eye with a combination of multiple small and medium drusen, although both may have similar values in total drusen area and volume. However, individual drusen are not routinely measured on SD-OCT because doing this by hand is extremely time consuming. While and few automated methods have been proposed for the segmentation of drusen regions, to our knowledge, a method to accurately differentiate and quantify each individual druse in SD-OCT scans has not yet been described.

We have described a method for the automatic segmentation of drusen regions including differentia-tion of each individual druse. Additionally, we believe or method may be superior in the differentiation of small drusen by addressing several unsolved challenges emerging from the prior work: (1) possible errors and inaccuracies in the RPE segmentation that smooth out drusen, (2) identification of drusen in regions where the inner and outer segments (IS/OS) have similar reflectivity as RPE, and (3) identification of drusen in regions where the COST line, IS/OS and RPE are difficult to differentiate from each other. Our method, which uses an accurate segmentation of the RPE and IS/OS boundaries, seems more sensitive in estimating small drusen than those using smoother estimation of the RPE layer, overcoming these challenges to some degree. The iRPE and IS boundaries can be differentiated in regions of low contrast by using a 3D segmentation approach.

translational vision science & technology

Identifying drusen presence in both iRPE and IS boundaries further increases the segmentation sensitivity in regions with possible inaccuracies in the RPE segmentation.

In order to evaluate our method accuracy estimating drusen regions, we compared with hand-drawn segmentations as the gold standard. Because hand-drawn drusen segmentations are reader-dependent and not perfect as a gold standard, we initially characterized the segmentation differences produced by two experts and at two different sessions on 512 B-scans obtained from four different eyes. Such characterization gave us a reference to compare with when evaluating the segmentations produced by our automated method. Expert variability was similar both within the same expert at different sessions and between the two experts (Table 1), with differences likely due to the fact that the boundaries of RPE layers are obscured by noise and the difficulty in creating a large number of manual outlines. The segmentations produced by our automated method showed substantial agreement with the average of the manual segmentations and comparable to the results obtained from different readers. Automated segmentation using a previously proposed method also produced comparable results in this limited dataset (Dataset A). We also evaluated our method's accuracy in a larger patient dataset (143 drusen-present B-scans obtained from 143 different eyes) in which we obtained manual segmentations from a third expert. We observed that differences between our automated method and the manual segmentations were slightly higher for this dataset, but still relatively low and within the limits established by the measured reader agreement. However, the previous method presented much larger differences with the manual segmentations for this larger dataset, and these differences were statistically significant ($P < 0.05$). Although this might be due to the segmenting approach of a reader (e.g., the reader constantly overestimating or underestimating drusen borders), it appears that our improved segmentation method produces more accurate and stable segmentations than the previous approach.

We evaluated the repeatability and reproducibility of a set of features extracted from segmented drusen using our method. We chose these features for this evaluation because they were previously shown to be promising as biomarkers for predicting AMD progression. Overall, the repeatability of the measurements was high at both the 512 × 128 and 200 × 200 scanning protocols. When evaluating the reproduc-ibility across the two protocols, the percentage differences in the features were low overall, indicating high reproducibility. However, the differences across protocols for average drusen slope and mean reflectivity inside the drusen regions were significant. Although these measurements seemed repeatable using the same protocol, the observed differences across protocols may be due to intrinsic properties resulting from the scanning pattern. Differences in drusen slope may be derived from A-scan and B-scan density differences between protocols, with higher values in the 512 × 128 protocol due to its lower B-scan density. While small drusen may be present in several B-scans using the higher density protocol, they may be present in only one B-scan using a lower scan density, which causes an overestimation of drusen slope in the vertical axis. Differences in mean intensity inside drusen regions may correspond to expected intrinsic acquisition differences among protocols and/or an effect of the low variability observed among the cases employed in this study (of less than 10%, see Table 2). It is possible that a larger dataset with drusen of a larger range of reflectivity properties may produce differences among protocols that are not significant. Nevertheless, it is important to note that although these two features seem repeatable using the same scanning protocol, they do not seem interchangeable across protocols a priori.

When comparing the repeatability of drusen area and volume features generated by our method to those produced by the Cirrus system, our methods presented higher repeatability than Cirrus in the 512 × 128 protocol. On the other hand, the Cirrus system presented excellent repeatability using the 200 × 200 protocol, with our methods also presenting high repeatability. Our method also showed higher reproducibility across protocols than the Cirrus system. Repeatability in Cirrus measurements has been previously analyzed for both protocols in prior publications, showing differences in similar ranges for the 200 × 200 protocol (7.5%–8.3%)[31] and even higher differences for the 512 × 128 protocol (14%–136%).[39] Repeatability using manual outlines in the 512 × 128 protocol has also been reported as an average of 16.64% and 7.82% for area and volume measurements,[39] respectively, which is within the ranges observed for the automated quantifications produced by our method. The lower repeatability of Cirrus measurements in the 512 × 128 protocol and its lower reproducibility across protocols may be due to a lower sensitivity in Cirrus measurements when using a scan pattern of lower vertical resolution. This effect

has also been shown previously, were higher values were reported for measurements derived from manual segmentations than those from Cirrus automated analysis,[39,40] and explains the intervals of maximum and minimum differences (200% and 0%, respectively, in Table 3) observed for Cirrus repeatability at the 512×128 resolution and for its reproducibility across protocols. While small drusen may be detected in one scan, they may be completely missed in a repeated scan (values of 0 area and volume), leading to such ranges. We observed that the native Cirrus software failed to detect any drusen in 6 of the 90 scans evaluated with the 512×128 scan pattern and in none of those evaluated with the 200×200 scan pattern. Although details about the Cirrus drusen segmentation remain undisclosed, this behavior seems to be linked to a trade-off maximizing specificity and the consequent reduction of sensitivity in the 512×128 pattern, and could be relative to the software version employed in this analysis (Cirrus HD-OCT Review Software, version 7.0.1.290). On the other hand, our method seems more sensitive when outlining drusen regions, which can be observed in higher drusen area and volume average values. An example with these characteristics, where differences in drusen area at 5 mm between protocols were also the highest for our method, is shown in Figure 6. Nevertheless, Cirrus automated drusen analysis has shown superior accuracy than its main commercial competitors,[40] and its ability detecting small drusen may be improved in software versions later than the one available to us in our work.

Our work has several limitations. Accuracy of the segmented outlines was compared with manual markings, in which experts invariably disagree on how they draw the margins of these lesions. Although intrareader and interreader segmentation variance was observed, which limited the quality of the comparisons, this variability between experts also emphasizes the utility of an automatic segmentation method, which would provide reproducible results. A second limitation is the reduced number of repeated SD-OCT scans acquired for the repeatability and reproducibility of drusen quantifications. Although all of the extracted features showed no statistically significant differences in repeated scans from the same protocol and most of them showed no significant differences in repeated scans across protocols, it is possible that a larger dataset may reveal such differences. On the other hand, the inclusion of a larger number of eyes with drusen of more variable characteristics may increase feature variability across

cases and reduce the importance of differences found across repeated scans.

In conclusion, our improved drusen segmentation method shows promising accuracy in automatically segmenting drusen regions, with the added value of being able to distinguish individual drusen in SD-OCT scans. The quantitative properties derived from the segmentations produced by our method show high repeatability in scans acquired using the two protocols tested. Such properties also show high reproducibility across protocols, with the exception of drusen slope and mean intensity, which show small, but significant differences between protocols that may be derived from the intrinsic properties of the different acquisition patterns.

Acknowledgements

Disclosure: **L. de Sisternes**, employed at Carl Zeiss Meditec, Inc.; **G. Jonna**, None; **M.A. Greven**, None; **Q. Chen**, None; **T. Leng**, None; **D.L. Rubin**, None

References

1. Klein R, Klein B, Jensen S, Meuer S. The five-year incidence and progression of age-related maculopathy: the Beaver Dam Eye Study. *Ophthalmology*. 1997;104:7–21.
2. Wong RW, Richa DC, Hahn P, et al. Iron toxicity as a potential factor in AMD. *Retina*. 2007;27:997–1003.
3. Friberg TR, Bilonick RA, Brennen PM. Analysis of the relationship between drusen size and drusen area in eyes with age-related macular

degeneration. *Ophthalmic Surg Lasers Imaging*. 2011;42:369–375.

4. Ferris FL, Davis MD, Clemons TE, et al. A simplified severity scale for age-related macular degeneration: AREDS Report No. 18. *Arch Ophthalmol*. 2005;123:1570–1574.

5. Friberg TR, Bilonick RA, Brennen PM. Is drusen area really so important? An assessment of risk of conversion to neovascular AMD based on computerized measurements of drusen. *Invest Ophthalmol Vis Sci*. 2012;53:1742–1751.

6. Smith RT, Sohrab MA, Pumariega NM, et al. Drusen analysis in a human-machine synergistic framework. *Arch Ophthalmol*. 2011;129:40–47.

7. Seddon JM, Reynolds R, Yu Y, Rosner B. Validation of a prediction algorithm for progression to advanced macular degeneration subtypes. *JAMA Ophthalmol*. 2013;131:448–455.

8. Khanifar AA, Koreishi AF, Izatt JA, Toth CA. Drusen ultrastructure imaging with spectral domain optical coherence tomography in age-related macular degeneration. *Ophthalmology*. 2008;115:1883–1890.

9. de Sisternes L, Simin N, Tibshirani R, et al. Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progression. *Invest Ophthalmol Vis Sci*. 2014;55: 7093–7103.

10. Huang D, Swanson EA, Lin CP, et al. Optical coherence tomography. *Science*. 1991;254:1178–1181.

11. Jain N, Farsiu S, Khanifar AA, et al. Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs. *Invest Ophthalmol Vis Sci*. 2010;51:4875–4883.

12. Bower BA, Chiu SJ, Davies E, et al. Development of quantitative diagnostic observables for age-related macular degeneration using Spectral Domain OCT. *Proc SPIE*. 2007;6426:64260W.

13. Freeman SR, Kozak I, Cheng L, et al. Optical coherence tomography-raster scanning and manual segmentation in determining drusen volume in age-related macular degeneration. *Retina*. 2010; 30:431–435.

14. Chen Q, Leng T, Zheng L, et al. Automated drusen segmentation and quantification in SD-OCT images. *Med Img Anal*. 2013;17:1058–1072.

15. Chiu SJ, Xiao TL, Nicholas P, et al. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt Express*. 2010;18:19413–19428.

16. de Sisternes L, Hu J, Rubin DL, Marmor F. Localization of damage in progressive hydroxy-

chloroquine retinopathy on and off the drug: inner versus outer retina, parafovea versus peripheral fovea. *Invest Ophthalmol Vis Sci*. 2015;56:3415–3426.

17. Ehnes A, Wenner Y, Friedburg C, et al. Optical coherence tomography (OCT) device independent intraretinal layer segmentation. *Transl Vis Sci Technol*. 2014;3(1):1–16.

18. Fernandez DC, Salinas HM, Puliafito CA. Automated detection of retinal layer structures on optical coherence tomography images. *Opt Express*. 2005;13:10200–10216.

19. Garvin MK, Abramoff MD, Wu X, et al. Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *Trans Med Imaging*. 2009; 28:1436–1477.

20. Ishikawa H, Jongsick K, Friberg TR, et al. Three-dimensional optical coherence tomography (3D-OCT) image enhancement with segmentation-free contour modeling C-mode. *Invest Ophthalmol Vis Sci*. 2009;50:1344–1349.

21. Kafieh R, Rabbani H, Abamoff MD, Sonka M. Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. *Med Image Anal*. 2013;17:907–928.

22. Li K, Wu X, Chen DZ, Sonka M. Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Trans Pattern Anal Mach Intell*. 2006;28:119–134.

23. Mishra A, Wong A, Bizheva K, Clausi DA. Intra-retinal layer segmentation in optical coherence tomography images. *Opt Express*. 2009;17:23719–23728.

24. Mujat M, Raymond CC, Cense B, et al. Retinal nerve fiber layer thickness map determined from optical coherence tomography images. *Opt Express*. 2005;13:9480–9491.

25. Toth CA, Farsiu S, Chiu SJ, et al. Automatic drusen segmentation and characterization in spectral domain optical coherence tomography (SDOCT) images of AMD eyes. *Invest Ophthalmol Vis Sci*. 2008;49:5394.

26. Farsiu S, Chiu SJ, Izatt JA, Toth CA. Fast detection and segmentation of drusen in retinal optical coherence tomography images. *Ophthalmic Tech XVIII, Proc SPIE*. 2008;68:20.

27. Chiu SJ, Izatt JA, O'Connell RV, et al. Validated Automatic segmentation of AMD pathology including drusen and geographic atrophy in SD-OCT images. *Invest Ophthalmol Vis Sci*. 2012;53: 53–61.

28. Farsiu S, Chiu SJ, O'Connell RV, et al. Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography. *Ophthalmology*. 2014;121:162–172.

29. Yi K, Mujat M, Park BH, et al. Spectral domain optical coherence tomography for quantitative evaluation of drusen and associated structural changes in non-neovascular age-related macular degeneration. *Br J Ophthalmol*. 2009;93:176–181.

30. Iwama D, Hangai M, Ooto S, et al. Automated assessment of drusen using three-dimensional spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2012;53:1576–1583.

31. Gregori G, Wang F, Rosenfeld PJ, et al. Spectral domain optical coherence tomography imaging of drusen in non-exudative age-related macular degeneration. *Ophthalmology*. 2011;118:1373–1379.

32. Bookstein FL. Principal warps: thin plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell*. 1989;11:567–585.

33. D'Errico J. 2010. Surface Fitting using gridfit. Available at: http://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit. Accessed: July 12, 2013.

34. Savitzky A, Golay MJE. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. 1964;36 8:1627–1639.

35. Meyer F. Topographic distance and watershed lines. *Signal Proc*.1994;38:113–125.

36. Pearson K. Notes on regression and inheritance in the case of two parents. *Proc Royal Soc Lond*. 1895;58:240–242.

37. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18:50–60.

38. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003;100:9440–9445.

39. Nittala MG, Ruiz-Garcia H, Sadda SR. Accuracy and reproducibility of automated drusen segmentation in eyes with non-neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2012;53:8319–8324.

40. Shlanitz FG, Ahlers C, Sacu S, et al. Performance of drusen detection by spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci*. 2010;51:6715–6721.
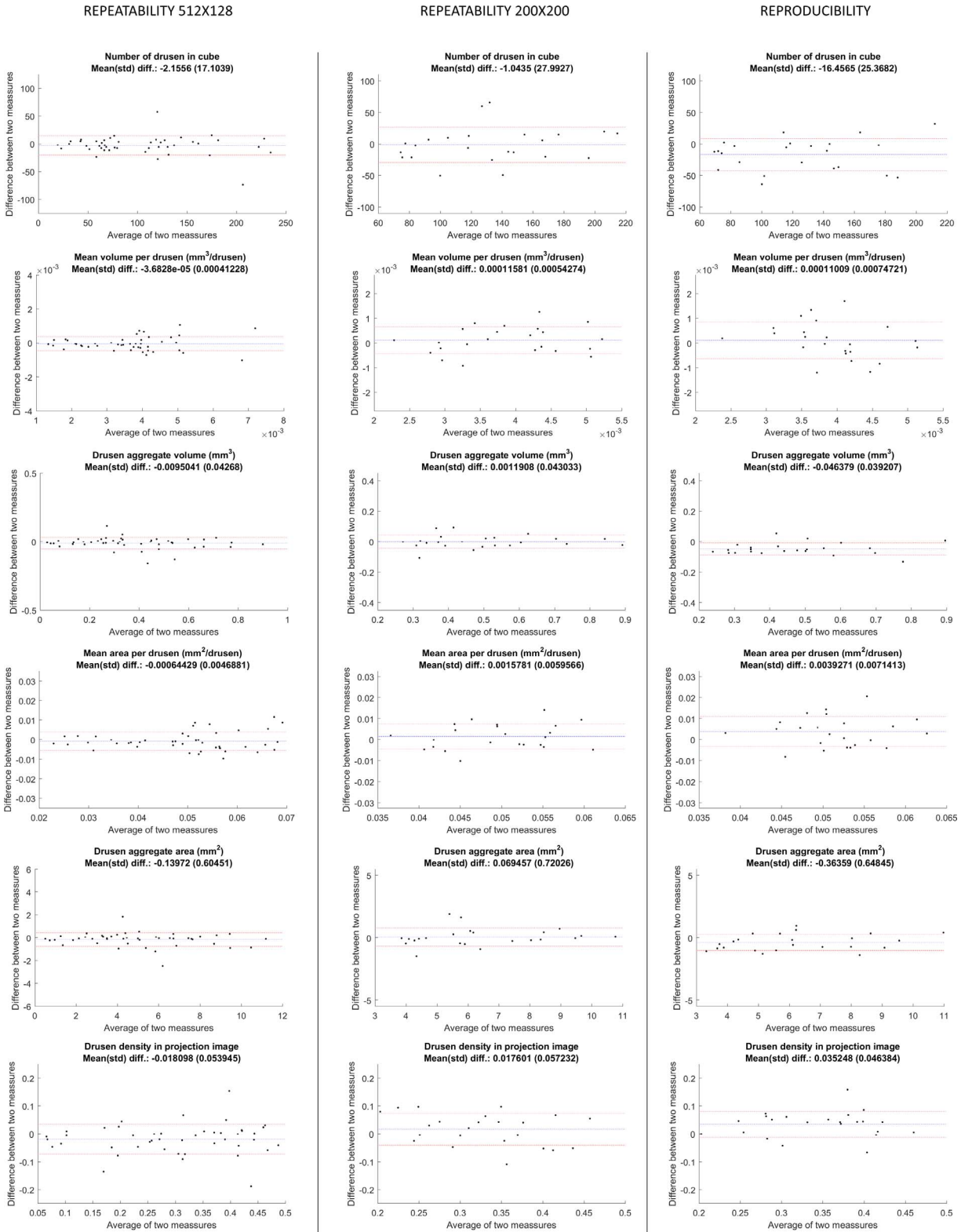
translational vision science & technology

Appendix

Figure 7. Bland-Altman plots for features investigated in Table 2 (continued in Fig. 8). *Left*, *middle*, and *right columns* correspond to differences measured in the 512 × 128 pattern, 200 × 200 pattern and across patterns, respectively.
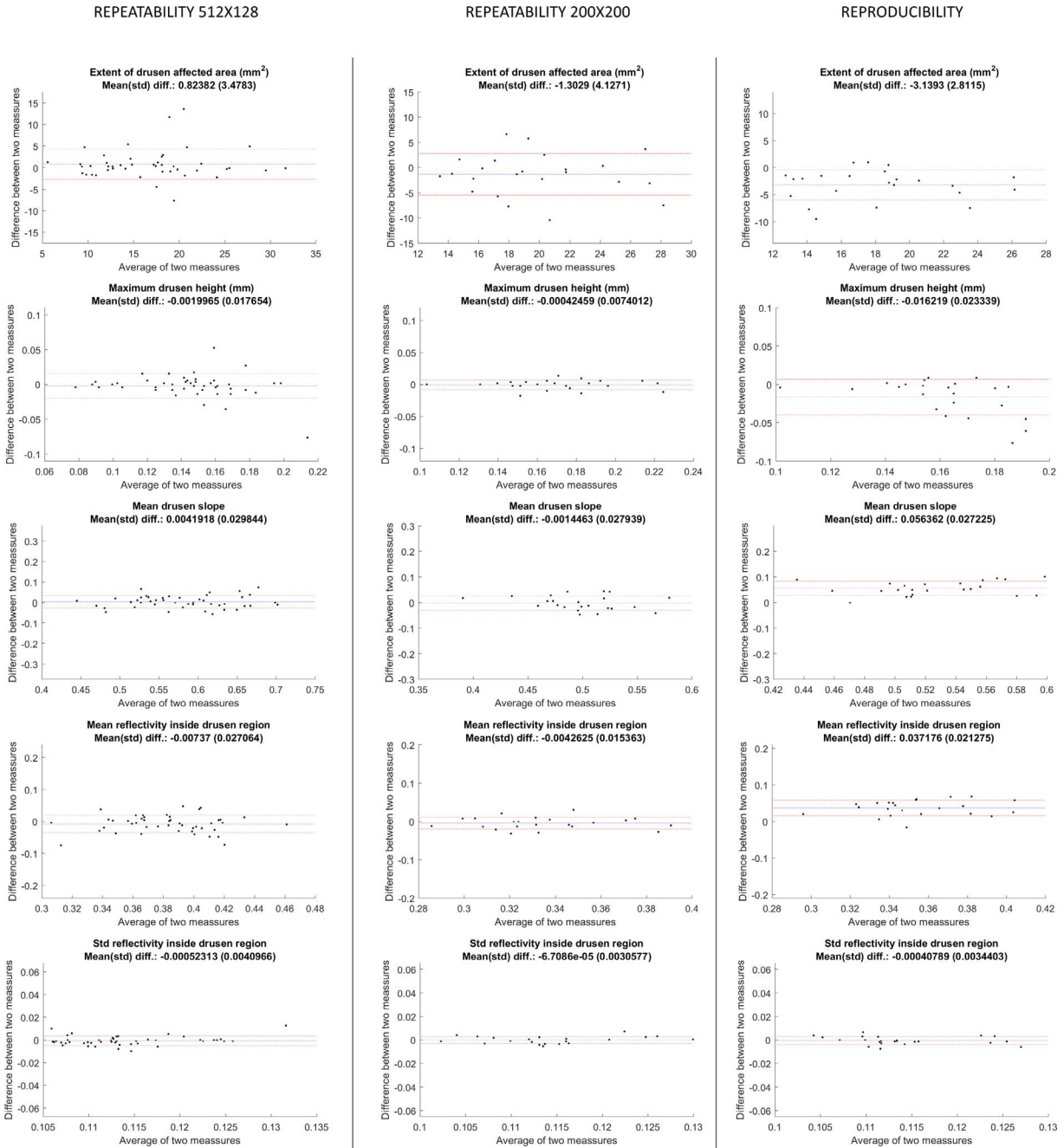
Figure 8. Bland-Altman plots for features investigated in Table 3 (continued from Fig. 7). *Left*, *middle*, and *right columns* correspond to differences measured in the 512 × 128 pattern, 200 × 200 pattern and across patterns, respectively.
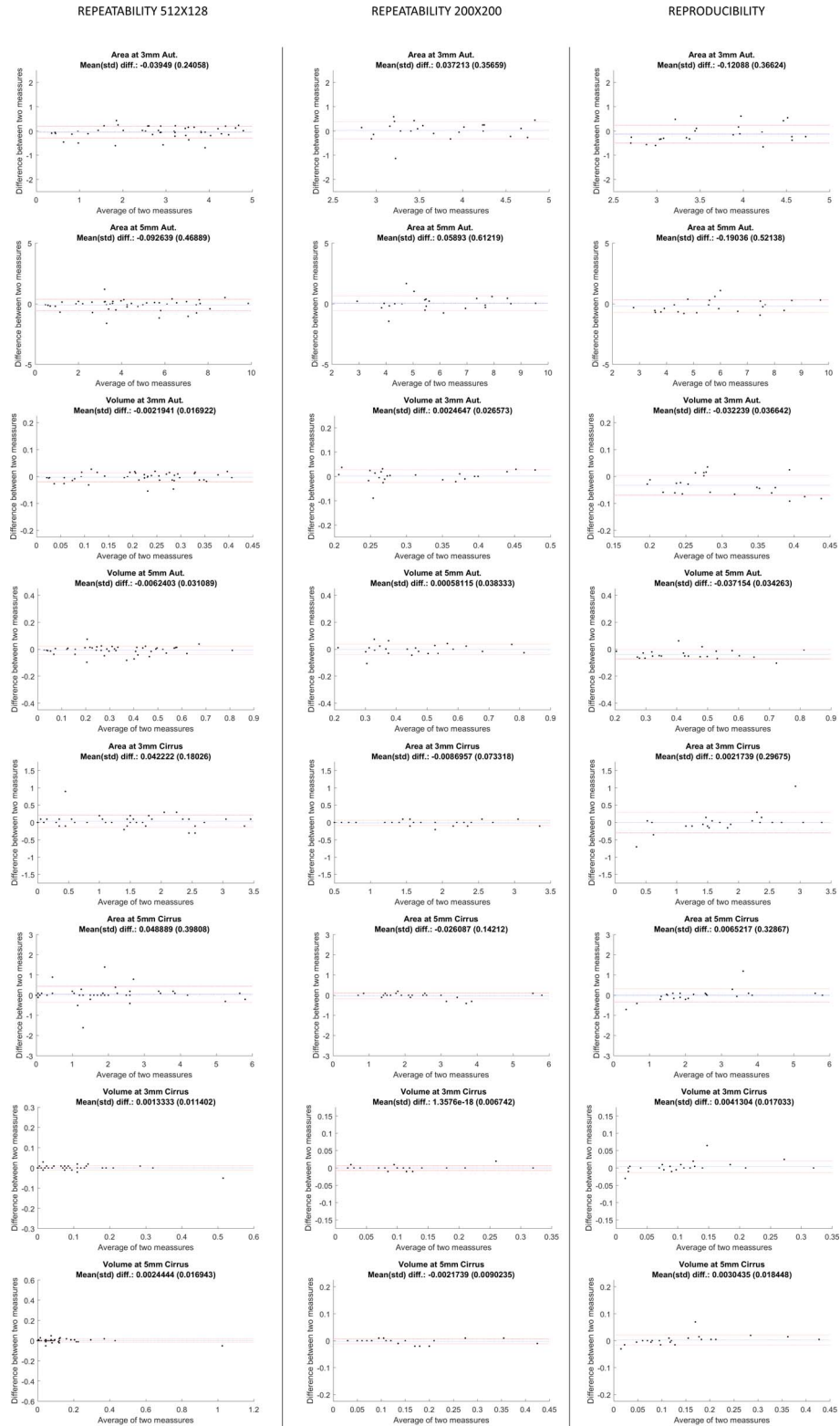
Figure 9. Bland-Altman plots for features investigated in Table 3. *Left, middle,* and *right columns* correspond to differences measured in the 512 × 128 pattern, 200 × 200 pattern and across patterns, respectively.