Retinal Nerve Fiber Layer Segmentation on FD-OCT Scans of Normal Subjects and Glaucoma Patients

Markus A. Mayer,[1,2] **Joachim Hornegger,**[1,2] **Christian Y. Mardin,**[3] **and Ralf P. Tornow**[3]

[1] *Pattern Recognition Lab, Martensstrasse 3, 91058 Erlangen, Germany*

[2] *Erlangen Graduate School in Advanced Optical Technologies (SAOT), Paul-Gordan Str. 6, 91052 Erlangen, Germany*

[3] *Department of Ophthalmology, Universitätsklinikum, Schwabachanlage 6, 91054 Erlangen, Germany*

markus.mayer@informatik.uni-erlangen.de

Abstract: Automated measurements of the retinal nerve fiber layer thickness on circular OCT B-Scans provide physicians additional parameters for glaucoma diagnosis. We propose a novel retinal nerve fiber layer segmentation algorithm for frequency domain data that can be applied on scans from both normal healthy subjects, as well as glaucoma patients, using the same set of parameters. In addition, the algorithm remains almost unaffected by image quality. The main part of the segmentation process is based on the minimization of an energy function consisting of gradient and local smoothing terms. A quantitative evaluation comparing the automated segmentation results to manually corrected segmentations from three reviewers is performed. A total of 72 scans from glaucoma patients and 132 scans from normal subjects, all from different persons, composed the database for the evaluation of the segmentation algorithm. A mean absolute error per A-Scan of 2.9 μm was achieved on glaucomatous eyes, and 3.6 μm on healthy eyes. The mean absolute segmentation error over all A-Scans lies below 10 μm on 95.1% of the images. Thus our approach provides a reliable tool for extracting diagnostic relevant parameters from OCT B-Scans for glaucoma diagnosis.

© 2010 Optical Society of America

OCIS codes: (110.4500) Optical coherence tomography

References and links

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C.A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," Science **254**(5035), 1178–1181 (1991).
2. B. E. Klein, R. Klein, W. E. Sponsel, T. Franke, L. B. Cantor, J. Martone, and M. J. Menage, "Prevalence of glaucoma. The Beaver Dam Eye Study," Ophthalmology **99**(10), 1499–1504 (1992).
3. H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," British Journal of Ophthalmology **90**, 262–267 (2006).
4. V. Guedes, J. S. Schuman, E. Hertzmark, G. Wollstein, A. Correnti, R. Mancini, D. Lederer, S. Voskanian, L. Velazquez, H. M. Pakter, T. Pedut-Kloizman, J. G. Fujimoto, and C. Mattox, "Optical coherence tomography measurement of macular and nerve fiber layer thickness in normal and glaucomatous human eyes," Ophthalmology **110**(1), 177–189 (2003).
5. G. Wollstein, J. S. Schuman, L. L. Price, A. Aydin, P. C. Stark, E. Hertzmark, E. Lai, H. Ishikawa, C. Mattox, J. G. Fujimoto, and L. A. Paunescu, "Optical Coherence Tomography Longitudinal Evaluation of Retinal Nerve Fiber Layer Thickness in Glaucoma," Archives of Ophthalmology **123**(4), 464–470 (2005).

6. V. Polo, J. M. Larrosa, A. Ferreras, F. Mayoral, V. Pueyo, and F. M. Honrubia, "Retinal nerve fiber layer evaluation in open-angle glaucoma. Optimum criteria for optical coherence tomography," Ophthalmologica **223**(1), 2–6 (2008).

7. F. K. Horn, C. Y. Mardin, R. Laemmer, D. Baleanu, A. Juenemann, F. E. Kruse, and R. P. Tornow, "Correlation between Local Glaucomatous Visual Field Defects and Loss of Nerve Fiber Layer Thickness measured with Scanning Laser Polarimetry and Spectral Domain Optical Coherence Tomograpy," Investigative Ophthalmology and Visual Science (IOVS) **50**(5), 1971–1977 (2009).

8. D. Koozekanani, K. Boyer, and C. Roberts, "Retinal thickness measurements from optical coherence tomography using a Markov boundary model," Medical Imaging, IEEE Transactions on **20**(9), 900–916 (2001).

9. J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in Optical Coherence Tomography," Journal of Biomedical Optics **4**(1), 95–105 (1999).

10. H. Ishikawa, S. Piette, J. M. Liebmann, and R. Ritch, "Detecting the inner and outer borders of the retinal nerve fiber layer using optical coherence tomography," Graefe's Archive for Clinical and Experimental Ophthalmology **240**(5), 362–371 (2002).

11. H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, "Macular Segmentation with Optical Coherence Tomography," Investigative Ophthalmology and Visual Science (IOVS) **46**(6), 2012–2017 (2005).

12. D. M. Stein, H. Ishikawa, R. Hariprasad, G. W. R. J., Noecker, J. G. Fujimoto, and J. S. Schuman, "A new quality assessment parameter for optical coherence tomography," British Journal of Ophthalmology **90**(2), 186–190 (2006).

13. M. Shahidi, Z. Wang, and R. Zelkha, "Quantitative Thickness Measurement of Retinal Layers Imaged by Optical Coherence Tomography," American Journal of Ophthalmology **139**(6), 1056 – 1061 (2005).

14. D. C. Fernández, H. M. Salinas, and C. A. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images," Optics Express **13**(25), 10200–10216 (2005).

15. G. Gilboa, N. A. Sochen, and Y. Y. Zeevi, "Image Enhancement and Denoising by Complex Diffusion Processes," IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 1020–1036 (2004).

16. M. Mujat, R. Chan, B. Cense, B. Park, C. Joo, T. Akkin, T. Chen, and J. de Boer, "Retinal nerve fiber layer thickness map determined from optical coherence tomography images," Optics Express **13**(23), 9480–9491 (2005).

17. G. M. Somfai, H. M. Salinas, C. A. Puliafito, and D. C. Fernández, "Evaluation of potential image acquisition pitfalls during optical coherence tomography and their influence on retinal image segmentation," Journal of Biomedical Optics **12**(4), 041209 (2007).

18. M. Baroni, P. Fortunato, and A. L. Torre, "Towards quantitative analysis of retinal features in optical coherence tomography," Medical Engineering & Physics **29**(4), 432–441 (2007).

19. O. Tan, G. Li, A. T.-H. Lu, R. Varma, D. Huang, and A. I. for Glaucoma Study Group, "Mapping of Macular Substructures with Optical Coherence Tomography for Glaucoma Diagnosis," Ophthalmology **115**(6), 949–956 (2008).

20. O. Tan, V. Chopra, A. T.-H. Lu, J. S. Schuman, H. Ishikawa, G. Wollstein, R. Varma, and D. Huang, "Detection of Macular Ganglion Cell Loss in Glaucoma by Fourier-Domain Optical Coherence Tomography," Ophthalmology **116**(12), 2305–2314 (2009).

21. M. Haeker, M. Abramoff, R. Kardon, and M. Sonka, "Segmentation of the Surfaces of the Retinal Layer from OCT Images," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006*, 800–807 (2006).

22. M. K. Garvin, M. D. Abramoff, R. Kardon, S. R. Russell, X. Wu, and M. Sonka, "Intraretinal Layer Segmentation of Macular Optical Coherence Tomography Images Using Optimal 3-D Graph Search," IEEE Transactions on Medical Imaging **27**(10), 1495–1505 (2008).

23. G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka, "Three-Dimensional Analysis of Retinal Layer Texture: Identification of Fluid-Filled Regions in SD-OCT of the Macula," Medical Imaging, IEEE Transactions on **29**(6), 1321 –1330 (2010).

24. D. Tolliver, I. Koutis, H. Ishikawa, J. S. Schuman, and G. L. Miller, "Unassisted Segmentation of Multiple Retinal Layers via Spectral Rounding," in *ARVO 2008 Annual Meeting* (Association for Research in Vision and Ophthalmology, Fort Lauderdale, 2008), Poster.

25. A. Mishra, A. Wong, K. Bizheva, and D. A. Clausi, "Intra-retinal layer segmentation in optical coherence tomography images," Optics Express **17**(26), 23719–23728 (2009).

26. A. Yazdanpanah, G. Hamarneh, B. Smith, and M. Sarunic, "Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach," in *MICCAI '09: Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 649–656 (Springer-Verlag, Berlin, Heidelberg, 2009).

27. V. Kajić, B. Považay, B. Hermann, B. Hofer, D. Marshall, P. L. Rosin, and W. Drexler, "Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis," Optics Express **18**(14), 14730–14744 (2010).

28. S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," Opt. Express **18**(18), 19413–19428

(2010).

29. K. Vermeer, J. van der Schoot, J. de Boer, and H. Lemij, "Automated Retinal and NFL Segmentation in OCT Volume Scans by Pixel Classification," in *ARVO 2010 Annual Meeting* (Association for Research in Vision and Ophthalmology, Fort Lauderdale, 2010), Poster.

30. E. Götzinger, M. Pircher, W. Geitzenauer, C. Ahlers, B. Baumann, S. Michels, U. Schmidt-Erfurth, and C. K. Hitzenberger, "Retinal pigment epithelium segmentation bypolarization sensitive optical coherencetomography," Optics Express **16**(21), 16410–16422 (2008).

31. A. R. Fuller, R. J. Zawadzki, S. Choi, D. F. Wiley, J. S. Werner, and B. Hamann, "Segmentation of Three-dimensional Retinal Image Data," IEEE Transactions on Visualization and Computer Graphics **13**(6), 1719–1726 (2007).

32. M. Szulmowski, M. Wojtkowski, B. Sikorski, T. Bajraszewski, V. J. Srinivasan, A. Szkulmowska, J. J. Kaluzny, J. G. Fujimoto, and A. Kowalczyk, "Analysis of posterior retinal layers in spectral optical coherence tomography images of the normal retina and retinal pathologies," Journal of Biomedical Optics **12**(4) (2007).

33. S. Joeres, J. W. Tsong, P. G. Updike, A. T. Collins, L. Dustin, A. C. Walsh, P. W. Romano, , and S. R. Sadda, "Reproducibility of Quantitative Optical Coherence Tomography Subanalysis in Neovascular Age-Related Macular Degeneration," Investigative Ophthalmology and Visual Science (IOVS) **48**(9), 4300–4307 (2007).

34. S. R. Sadda, S. Joeres, Z. Wu, P. Updike, P. Romano, A. T. Collins, and A. C. Walsh, "Error Correction and Quantitative Subanalysis of Optical Coherence Tomography Data Using Computer-Assisted Grading," Investigative Ophthalmology and Visual Science (IOVS) **48**(2), 839–848 (2007).

35. D. Baleanu, R. P. Tornow, F. K. Horn, R. Laemmer, C. W. Roessler, A. G. Juenemann, F. E. Kruse, and C. Y. Mardin, "Retinal Nerve Fiber Layer Thickness in Normals Measured by Spectral Domain OCT," Journal of Glaucoma **19**(7), 475–482 (2009).

36. C. R. Vogel and M. E. Oman, "Iterative Methods for Total Variation Denoising," SIAM Journal on Scientific Computing **17**(1), 227–238 (1996).

37. T. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," SIAM journal on numerical analysis **36**(2), 354–367 (1999).

38. A. Fercher, W. Drexler, C. Hitzenberger, and T. Lasser, "Optical coherence tomography - principles and applications," Reports on Progress in Physics **66**(2), 239–303 (2003).

39. D. C. Hood, B. Fortune, S. N. Arthur, D. Xing, J. A. Salant, R. Ritch, and J. M. Liebmann, "Blood vessel contributions to retinal nerve fiber layer thickness profiles measured with optical coherence tomography," Journal of Glaucoma **17**(7), 519–528 (2008).

40. D. C. Hood, A. S. Raza, K. Y. Kay, S. F. Sandler, D. Xin, R. Ritch, and J. M. Liebmann, "A comparison of retinal nerve fiber layer (RNFL) thickness obtained with frequency and time domain optical coherence tomography (OCT)," Optics Express **17**(5), 3997–4003 (2009).

1. Introduction

Ophthalmology has been one of the main application areas of Optical Coherence Tomography (OCT) [1] since its invention in 1991. OCT allows a direct visualization of the retina and its layered structure. This proved to be very beneficial for glaucoma disease research. Glaucoma is one of the most frequent reasons for blindness in the world [2, 3]. During glaucoma progression the supporting tissue and nerve fibers in the retina are lost. Thus a thinning of the innermost retinal layer, the retinal nerve fiber layer (RNFL), is observed [4, 5, 6, 7]. Since the first appearance of commercially available OCT systems, automated retinal layer segmentation algorithms were presented to objectively quantify the RNFL thickness and its loss. While segmentation algorithms are built into commercial systems, details of their design remain undisclosed. Below we give a short overview of only the published research on retinal layer segmentation.

To our knowledge, Koozekanani et al. [8] presented for the first time an automated retina segmentation algorithm on Time Domain (TD) OCT scans. This first publication on the topic already mentioned one main challenge. The noise that corrupts OCT images is non-Gaussian, multiplicative and neighborhood correlated (see also Schmitt et al. [9]). Thus, it can not be easily suppressed by standard software denoising methods. Besides this observation one key element of subsequent algorithms was already included in this early paper: An edge detection approach was introduced which also takes the leading sign of the derivative into account so as to differentiate between rising and falling contrast along a depth profile [8]. This method outperformes simple thresholding algorithms, as they tend to fail due to the high variance of intensity

values on OCT images. Their evaluation considered only healthy normal subjects. Ishikawa et al. [10] included glaucoma patients in their performance analysis. To our knowledge, this was the first RNFL segmentation algorithm published and the only group that evaluates the algorithm on both glaucoma patients and normal subjects. In a subsequent paper five different layers were segmented [11]. For the evaluation the quality of the images was measured with the quality index of Stein et al. [12]. It was stated that the automated segmentation fails on more than half of the low quality images. According to Ishikawa et al. a good quality scan is harder to achieve on diseased eyes, which makes the automated segmentation challenging. Additionally, a visualization of the RNFL thickness in 2D interpolated out of multiple radial scans was shown. While Ishikava et al. used an integrity check to connect neighboring 1D depth profile segmentations, Shahidi et al. [13] just averaged thickness profiles in transversal direction. This lead to thickness profiles of much coarser resolution. Fernandez et al. [14] proposed complex diffusion (see Gilboa et al. [15]) instead of the typical median filtering as a preprocessing step for the segmentation. Seven retinal layers including the inner/outer nuclear layer and inner/outer plexiform layer were segmented. It is stated that pathologies may violate assumptions made in the algorithm and thus parameters have to be adjusted. In the same year Mujat et al. [16] presented an algorithm to segment the RNFL on Frequency Domain (FD) OCT Volume data. The segmentation was done frame-by-frame in 2D. While results were shown only from two volumes of healthy normal subjects and no evaluation was performed, the number of steps included in the algorithm leads to the assumption that FD data, although of higher resolution and with fewer motion artifacts, presents more challenges to segmentation developers. Somfai et al. [17] investigated how decreased image quality caused by operator errors affects segmentation results.

Baroni et al. [18] formulated an edge-likelihood function consisting of a gradient and a smoothness term to segment TD-OCT images. The most recent work still concentrating on the segmentation of TD-OCT images was published by Tan et al. [19]. A segmentation based on progressive edge detection was presented that regularizes the segmentation result by averaging fewer A-Scans in each segmentation refinement step. The algorithm itself was not evaluated, since the scope of the work was not on the algorithm development, but on the investigation of parameters generated from the segmentations for glaucoma diagnosis. Images with segmentation errors were discarded from the presented study. Later Tan et al. investigated further glaucoma diagnostic parameters on 3D FD-OCT volume scans [20].

The first method that really made use of 3D information is the retina segmentation from Haecker et al. [21] which was further developed to a multilayer segmentation by Garvin et al. [22]. Here six radial linear TD-OCT 2D scans were combined to a volume. A 3D-Graph search to minimize a cost function for each layer (for up to five layers) was performed to segment the volume. For each boundary, assumptions in the cost function were made that could include e.g. a signed edge, summation of pixel intensities in limited regions and summation of pixel intensities from already segmented borders. The algorithm was evaluated on data from subjects with unilateral chronic anterior ischemic optic neuropathy. No glaucoma patients were included in the evaluation. The approach was further extended to all 10 retinal layers in Quellec et al. [23]. In this work, the focus shifted from the actual segmentation of the retinal layers to an application of the segmentation results. Thickness measurements, together with texture features from image data within the layer boundaries, are used to build up a model for the detection of pathologic retina abnormalities like fluid filled regions.

Recently, multiple sophisticated methods known from computer science and optimization theory were applied and modified for OCT-Data. Tolliver et al. [24] uses spectral rounding to detect the RNFL on FD-OCT data of normal subjects and patients. Mishra et. al. [25] uses a two step optimization scheme solved with dynamic programing to segment all retinal layers.

It was applied on scans of rat eyes, not on data from humans. The same holds for the work of Yazdanpanah et al. [26], who uses active contours. An enery functional that includes a shape prior is minimized. Such prior information can be either formulated out of heuristic considerations, or computed from training data. Kajić et al. [27] use the latter in their appearance shape model based approach. The evaluation shows that the algorithm is very robust and insensitive to noise. However the results were only generated from data of normal eyes. The evaluation from Chiu et al. [28] is also performed only on data of normal subjects, but visual examples in the paper show that plausible segmentations even in severe pathologic cases can be generated by the method. It is based on graph theory and dynamic programming. Vermeer et al. [29] also avoid heuristics in their approach. A support vector machine is trained to classify every pixel of a volume for its position above or below certain boundaries. The resulting segmentation is regularized by a level set method. The classifier was trained on data of normal subjects. An evaluation was performed on volume scans of 10 normal subjects and 8 glaucomatous patient, but only one or two B-Scans per volume scan were taken into account for the evaluation. The segmentation algorithm showed decreased performance on the pathological data.

The work of Götzinger et al. [30] used the additional information of a polarization sensitive OCT system provides for segmenting the retinal pigment epithelium with two different methods. Patient data was shown, but a quantitative evaluation was not performed.

Contrary to the aforementioned boundary search approaches, Fuller et al. [31] proposed a half-manual region-based classification to segment layers. Out of manually marked regions on sample frames, a SVM classifier was trained for segmenting the whole FD-volume scan. A similar classifier-based half-manual approach was followed by Szulmowski et al. [32]. Joeres et al. [33] and Sadda et al. [34] presented a completely manual segmentation tool for TD-OCT Scans. In pathologies like age related macula degeneration, or pigment epithelial detachment automated algorithms will most likely fail due to heavy abnormalities in the data. They showed that a manual segmentation provides reliable and reproducible results in these cases.

In this paper we present a completely automated segmentation of the RNFL. To our knowledge this is the most important layer for glaucoma diagnosis and is thus the focus of our work. The segmentation is a 2D approach working on circular FD-OCT scans, but can be easily applied on 3D volume data, as will be shown. The goal during the development of the algorithm was to make as few assumptions on the layer borders as possible. The employed assumptions should also be very general. Reliable application on pathological cases should be possible without changing any parameter. In Section 2 our method is presented. Qualitative visual examples and quantitative evaluation results are shown in Section 3. Section 4 includes a summary, conclusions, as well as further ideas.

2. Method

We use two datasets that are described in the following subsection (2.1). The processing steps to segment the RNFL on circular B-Scans are covered in the subsequent two subsections. First the inner and outer retinal borders are detected (2.2). Particularly, the inner border of the retina is detected at the inner border of the RNFL, or internal limiting membrane (ILM). The outer retinal border is detected at the the outer retinal pigment epithelium boundary (RPE). The algorithm for finding the position of the outer nerve fiber layer boundary (ONFL) is described in (2.3). The method is then extended to 3D (2.4). The final subsection presents our evaluation methodology (2.5).

2.1. Data

Circular B-Scans were acquired from 204 subjects with a Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany). This OCT device is referred to as Spectralis for the re-
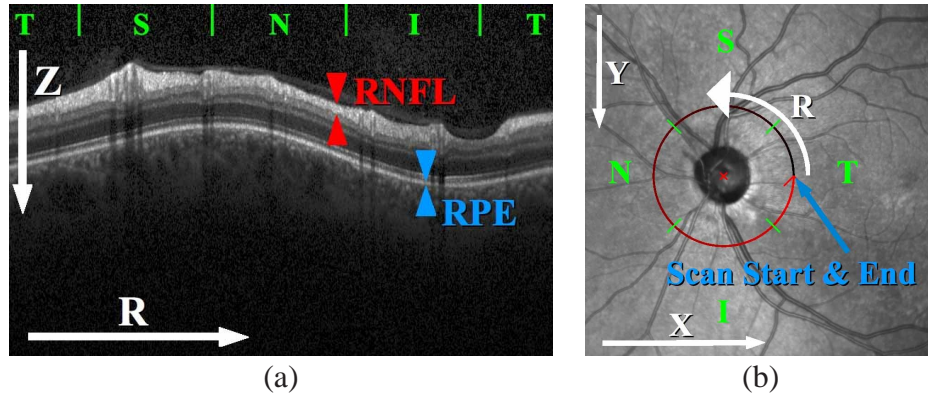
(a)	(b)

Fig. 1. Example circular B-Scan of a left eye with coordinate system denominations. Right eye denominations and scan pattern are equivalent and follow the common rules for the mapping between left and right eye. (a) OCT B-Scan. The retinal layers relevant for this work are marked: The retinal nerve fiber layer (RNFL) and retinal pigment epithelium (RPE). (b) SLO image captured by the Spectralis HRA+OCT during the same scanning process. The circular scan pattern position and its direction corresponding to the R-direction in the images is marked. The quadrant borders on the SLO image scan position and on the OCT scan are shown with green lines. The quadrants are: Temporal (T), Superior (S), Nasal (N), Inferior (I).

mainder of the paper. The scans were centered at the optic disk and had a diameter of 3.4mm. The B-Scans consisted of 512 or 768 A-Scans. Each A-Scan consists of 496 pixels. The axial resolution of the Spectralis is 7μm in tissue, although the pixel length is 3.87μm. The images are thus oversampled in the axial direction. The raw data was exported using the VOL file format of Heidelberg Engineering. The pixel intensity value range in the VOL files is [0;1], saved as 32 bit floating point values. All computations were performed in the same data format.

To clarify denominations in this work: A-Scan and depth profile are used interchangeably. B-Scan, OCT image and 2D frame are also used as synonyms. OCT volumes are also referred as 3D data. In the description of the algorithm the axial direction is Z. To simplify formulas, the transversal direction in a circular Scan, giving the position of an A-Scan in the resulting image, is denominated only by R. The transversal directions in a volume are X and Y. The Z direction as well as the Y direction have their origins in the upper left corner of the corresponding images. Figure 1 illustrates these notations. Unless stated otherwise, the intensity values of the VOL-File are double square routed for display reasons as proposed by Heidelberg Engineering. All abbreviations and symbols are shown in Table 5.

The recorded data has a wide range of subjectively perceived quality. Unfortunately, at the time the data was recorded the Spectralis software had no built-in quality index. Therefore we chose the following quality measure that, by visual inspection, correlates well with the subjectively perceived quality.

$$QI = 1 - \frac{\#N_{I(z,r)=0}}{\#N}; \qquad (1)$$

QI is the quality index (QI). $I(z,r)$ denotes the image intensity at position z in Z-direction and r in R-Direction. $\#N_{I(z,r)=0}$ is the number of pixels on the scan with intensity value 0. This is normalized by $\#N$, the complete number of pixels on the scan. The quality index is motivated as follows: The scans result from an averaging of multiple single scans. Fewer scans yield a higher

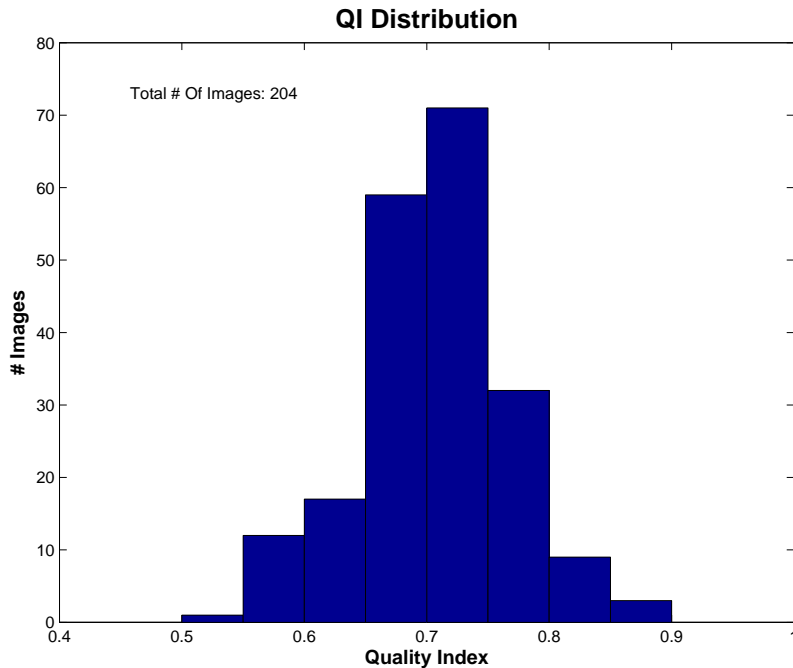**QI Distribution**

Total # Of Images: 204

Fig. 2. Quality index (QI) distribution of all 204 circular B-Scans in our dataset. A high QI value denotes a good quality image. 93.6% of the images have a QI above 0.6.

noise level. We observed that on the Spectralis images with high noise level some pixels have 0 intensity even inside the tissue. This is even more pronounced in the dark regions outside the retina. By averaging more images this effect diminishes, as the number of randomly distributed zero intensities significantly decreases. Because it is not possible to read out the number of single captures out of the VOL files in the present software version, our simple QI is presumed to be an objective measurement of the quality of an image. Subtracting the fraction from 1 yields an ascending index. High QI values correspond to a high quality image. The QI is not necessary in complete agreement with with a subjective assessment, but it provides a rough estimate for evaluation and data description. The distribution of QI in our circular B-Scan evaluation data set can be seen in Fig. 2. No image of the dataset was excluded due to quality reasons.

Two groups of subjects formed our circular B-Scan dataset: Glaucoma patients (72 subjects, age 62.1 ± 9.4, mean defect 5.5 ± 6.2) and healthy normal subjects (132 subjects, age 49.8 ± 16.3, mean defect 0.5 ± 0.6). All patients were members of the 'Erlangen Glaucoma Registry' with annual visits to our glaucoma service. The inclusion/exclusion criteria and the type of examinations are defined in a protocol which was approved by the Local Ethics committee. The study is registered at www.clinicaltrials.gov (NCT00494923). The study followed the tenets of the declaration of Helsinki for research involving human subjects and informed consent was obtained from all participants of the study.

Only one eye of each subject was taken into account for the evaluation. The subjects were diagnosed based on an ophthalmic examination using slit lamp inspection, applanation tonometry, funduscopy, gonioscopy, perimetry and papillometry. A 24 hours intraocular pressure profile with 6 determinations was also obtained. A detailed review of the employed diagnostic routine can be found in Baleanu et al. [35] and Horn et al. [7] and is not within the scope of this
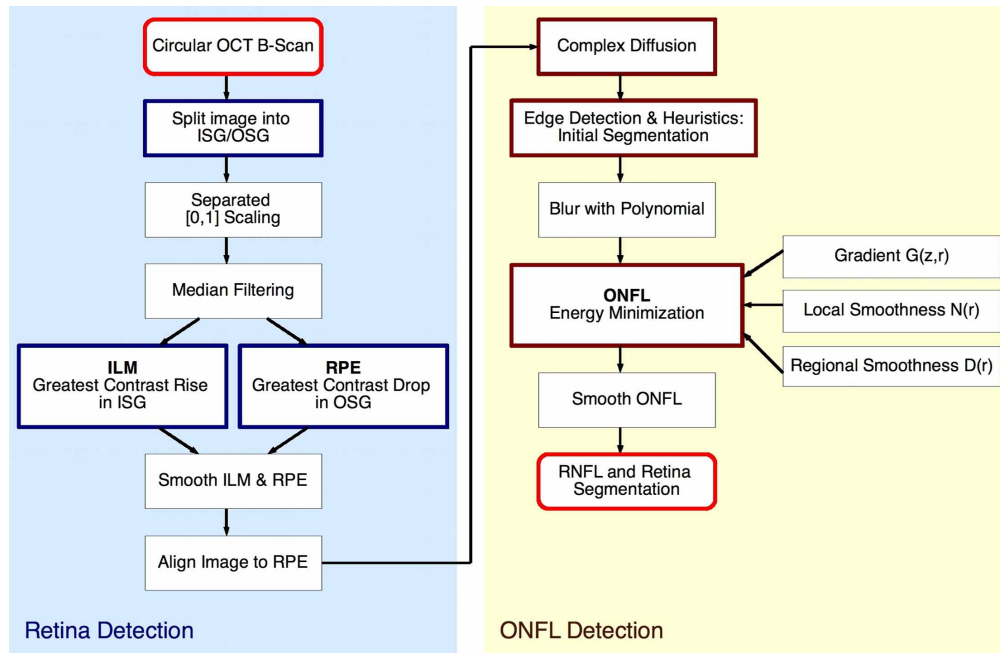
Fig. 3. Algorithm overview. Input and output data are marked in red. The retina detection is colored in blue, the outer nerve fiber layer detection in yellow. Important steps are marked with bold rectangles.

paper.

In addition to the circular B-Scans volume scans were acquired centered on the optic nerve head. The number of pixels are 512 in the X-direction, 97 in the Y-direction and 496 in the Z-direction. The pixel spacing in the X-direction is $11.55\mu m$, in the Y-direction $61.51\mu m$ and $3.87\mu m$ in the Z-direction. One volume scan from a glaucoma patient and one volume scan from a healthy normal subject are discussed in Section 3 as exemplary cases.

2.2. Detecting the retinal boundaries

All algorithm parameters were adapted by visually inspecting random sets of images from the database. The processing steps of the algorithm are shown in Fig. 3 with visual examples provided in Fig. 4. A scan from a glaucoma patient was chosen as an example. It shows a nearly complete loss of the RNFL in the transition between the inferior and temporal quadrant, while the other regions still have relative high RNFL thickness.

To limit the search space for the retinal boundaries, first a separating line located inside the outer nuclear layer is identified. It splits the image content into the the inner segment (ISG) and the outer segment (OSG) of the retina. The image is blurred with a wide Gaussian filter (standard deviation $\sigma = 22 pixels$). The separating line is the lowest intensity value inbetween the two maximas with the highest intensity value (see Fig. 4 (a) and Fig. 5(a)). The intensities of each A-Scan were scaled to $[0;1]$ in the ISG and OSG separately. For a rough speckle noise removal, a 2D median filter of size 5 in the Z- and 7 in the R-direction is applied twice, as proposed by Ishikawa et al. [11]. The ILM is then set to the greatest contrast rise in the ISG, while the RPE is set to the greatest contrast drop in the OSG. Both lines are smoothed. Smoothing includes an outlier detection by fitting a polynomial of degree 5 and removing distant line
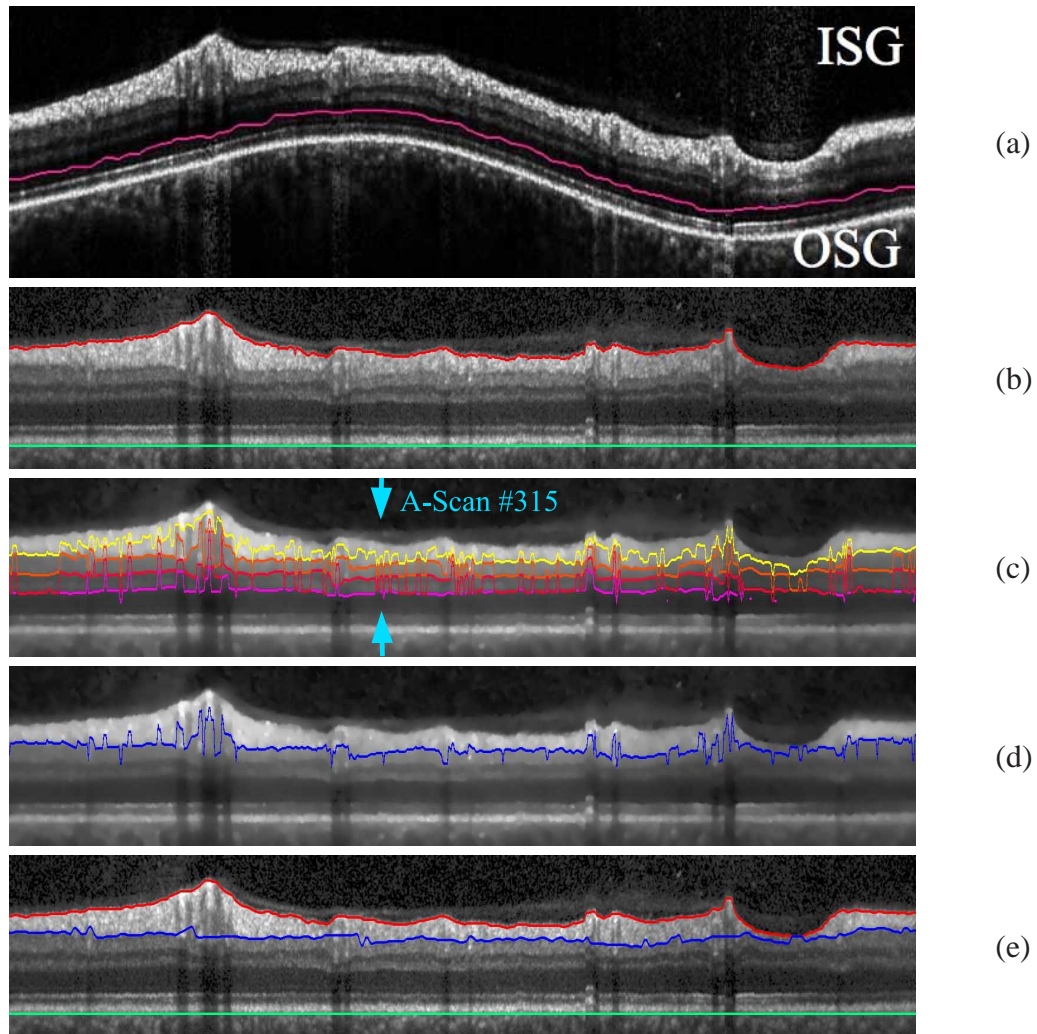
Fig. 4. Processing steps of the nerve fiber layer segmentation shown on an example scan of a glaucomatous eye (QI = 0.74) with local nerve fiber layer loss. (a) Separating line in the outer nuclear layer detected. Inner and outer segment of the retina are separately [0 : 1] scaled. ISG: Inner segment of the retina. OSG: Outer segment of the retina. (b) Inner nerve fiber layer boundary and retinal pigment epithelium detected. A-Scans aligned so that the retinal pigment epithelium forms a constant even line. The image intensities are changed back to the original ones. (c) Image denoised by complex diffusion. A maximum of four greatest contrast drops in the inner segment of the retina is detected. (d) Initial segmentation of the outer nerve fiber layer boundary formed by heuristic decisions. (e) Result after energy-minimization segmentation described in Section 2.3. The resulting mean RNFL thickness is $73.5\mu m$.
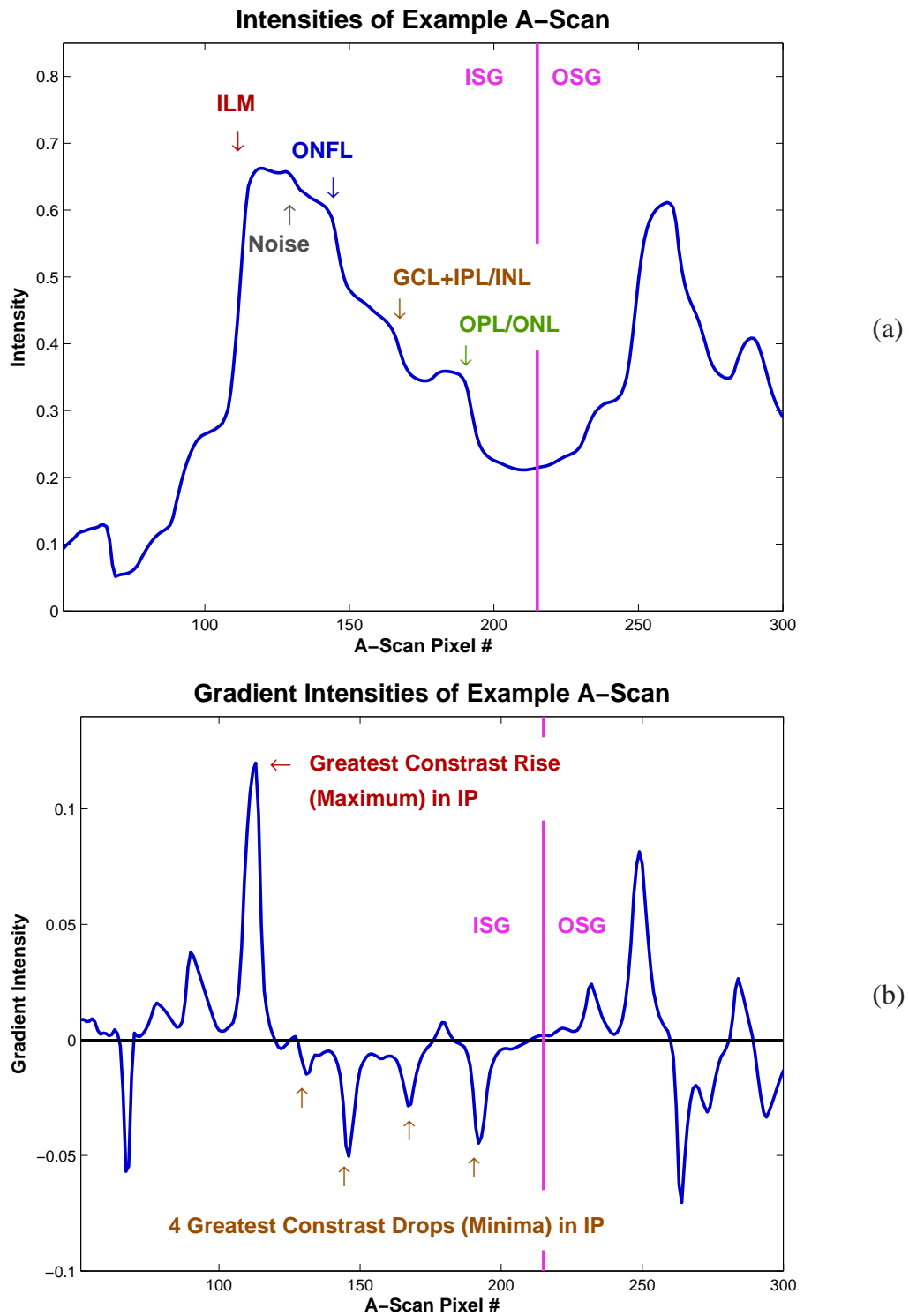
Fig. 5. (a) Intensity plot along an A-Scan and (b) its corresponding derivative. The A-scan # 315 of the denoised example image in Fig. 4 (c) is shown. It is cropped to the retina region. The intensity rise at the ILM, as well as the intensity drops at the ONFL, at the inner plexiform layer (IPL)/inner nuclear layer (INL) border and the outer plexiform layer

segments afterwards. Short disconnected segments are also removed, a median filter and Gaussian smoothing are applied. Gaps are closed by linear interpolation. After the smoothing, the A-Scans of the original unprocessed image are aligned so that the RPE forms a constant even line (see Fig. 4 (b)). This image becomes the basis for the following ONFL segmentation.

2.3. Detecting the outer nerve fiber layer boundary

Our empirical analysis of the OCT data showed that a simple edge detection for the ONFL, even if the retina boundaries are known, will not give promising results. This holds especially true for a general low image quality, glaucoma patients having a complete local loss of the RNFL and normal subjects with a very thick RNFL. For the last two cases, a state-of-the-art preprocessing with sophisticated denoising as proposed by Fernandez et al. [14] and Mujat et al [16] is also insufficient. A neighborhood integrity check as mentioned in [11] might not be able to cope with a jump of the segmented border in a whole region to a higher contrast outer layer border. Assumptions on the layers, as Garvin et al. [22] made, may be violated in pathological cases, or parameters have to be adapted for either normal subjects, or glaucoma patients. Our approach is the following: The image is denoised with complex diffusion (see Gilboa et al. [15]) as proposed by Fernandez et al. [14]. Our implementation is not based on the traditional time-marching implementation but uses lagged diffusivity [36, 37]. The code of the algorithm can be downloaded from the homepage of our work group (http://www5.informatik.uni-erlangen.de) on the personal page of the first author. The timestep parameter was set to 13, while the σ_{CD} parameter that controls which gradients are detected as edges, is directly estimated from the image:

$$\sigma_{CD} = \frac{1}{3} std(I - I_{medfilt}); \tag{2}$$

I denotes the original image matrix, $I_{medfilt}$ is the original image on which every A-Scan is filtered with a median filter of width 7. The $\frac{1}{3}$ is a heuristic weighting factor. The computation of the standard deviation of all pixels is abreviated by $std(...)$. The noise estimate does not correspond to a physically meaningful noise measurement on the OCT data, but it has by visual inspection proven to adapt to the different noise levels and qualities of the OCT B-Scans. After denoising the image, along the A-Scans the four greatest contrast drops are detected in the ISG (see Fig. 4 (c)) and Fig. 5(b)). Actually, only three layer borders with falling contrast should lie in the ISG beneath the ILM, namely: the ONFL; the border between the inner plexiform layer and the inner nuclear layer; the border between the outer plexiform layer and the outer nuclear layer. To derive an initial segmentation from the detected four minima, the following measurement is used:

$$PS(r) = \sum_{z \in ISG} I(z, r) / \sum_{z \in OSG} I(z, r) \tag{3}$$

The part sum ratio PS yields for every column position r on the B-Scan the ratio of the summed intensities in the ISG and OSG region. The heuristic rule to form an initial segmentation is: If PS is high ($PS > 0.7$), the second contrast drop beneath the ILM is chosen. This can be motivated by the idea, that a high PS indicates a thick RNFL. Thus, the first contrast drop is most likely speckle noise. In the other cases ($PS \leq 0.7$) the first contrast drop beneath the ILM is used. This does not hold if fewer than three contrast drops are detected in an A-Scan. A complete loss of the RNFL can be assumed in that case. The initial segmentation is set to the IFNL. This heuristic method delivers a good estimate of the segmentation (see Fig. 4 (d)), but can be further enhanced.

To improve the segmentation we formulate an energy-minimization based approach:

$$E(r) = G(z, r) + \alpha N(r) + \beta D(r); \tag{4}$$

$$\overset{min}{z} \, (E(r)) \forall r \Rightarrow ONFL(r). \tag{5}$$

$ONFL(r)$ gives the Z-position of the boundary at an A-Scan r. $E(r)$ is the energy at an A-Scan r that needs to be minimized. It consists of three terms. Two factors, α and β weight these terms. In the current implementation they are set to $\frac{1}{3000}$ and $\frac{2}{3000}$ respectively. The first term, $G(z,r)$ is the gradient at depth z. As the ONFL is in Z-direction a contrast fall-off, the gradient should have a negative sign with an absolute value which is as high as possible. $N(r)$ is a smoothing term that ensures that there are no high jumps in the border, while allowing for some edges. It is defined as the sum of the absolute differences in height z of the border in A-Scan r and its neighbors:

$$N(r) = |ONFL(r-1) - ONFL(r)| + |ONFL(r+1) - ONFL(r)|. \tag{6}$$

The second smoothness term $D(r)$ works on a more global scale. It is motivated by the observation, that when the A-Scans are aligned for an even RPE, the ONFL is partially almost even, too. In Baroni et al. [18] the distance to a constant line along the whole B-Scan was taken as a smoothness term. We extend this idea. The RNFL should not be as constant as possible over the whole B-Scan but within certain regions. To avoid using arbitrary positions on the image, the regions in between blood vessels are used. $D(r)$ is therefore the distance to the average height of the segmented boundary in between two blood vessels:

$$D(r) = ONFL(r) - (\sum_{r \in BVR} ONFL(r))/\#N_{r \in BVR} \tag{7}$$

BVR means the region in between two blood vessels, $\#N_{r \in BVR}$ is the number of A-Scans in this region. The blood vessel positions are determined by adaptive thresholding. A layer of 8 pixels above the RPE is summed along Z-direction to form a RPE intensity profile. The average of this profile is computed in a 61 pixel wide window. If the value of the profile in the middle of the window is lower than 0.7 times the average, it is marked as a blood vessel. For the BVR boundaries the centers of the blood vessels are computed. As the size parameter of the average window and the threshold are fixed, some large vessels above 12 pixels in width are not detected. This does not affect the segmentation results as typically enough blood vessel centers are detected for splitting the image into regions. Very small blood vessels with a diameter below 4 pixels are ignored in this computation.

The energy Equation (4) is solved iteratively by moving the boundaries in the directions of decreased energy. The initial segmentation is created out of high-contrast edges. As we want to avoid being stuck in this first estimate in the iterative process the initial segmentation is heavily blurred by fitting a polynomial of degree 4 trough it. This polynomial provides the initial values for the energy minimization process. The resulting boundary solution is smoothed similar to the RPE in Section 2.2. The result of the algorithm is shown in Fig. 4 (e).

2.4. 3D Application

The aforementioned method is applicable on 2D circular Scans. A direct application on 2D linear scans or slices from volumes leads to segmentation errors. The assumption that the ONFL is roughly piecewise constant is violated. Furthermore obvious splitting points, like the blood vessels, might be missing in some areas as for example the macula region. But a segmentation of 3D data can be achieved by interpolating circular B-Scans with varying diameters d out of the volumes. Given linear scans, circular scans are created through bilinear interpolation of the intensity information in the R-direction out of the X/Y aligned volumes. The center of the circular scans has to be set manually. The segmentation algorithm is then applied to the interpolated scans without any parameter change. A method for visualizing the results is computing a

RNFL thickness profile $RNFL_d(r)$ from the segmentation by calculating the distance between the ILM and ONFL:

$$RNFL_d(r) = (ONFL_d(r) - ILM_d(r)) * Scale_Z \qquad (8)$$

where $ONFL_d(r)$ and $ILM_d(r)$ denote the border positions in pixels at an A-Scan r for a scan circle diameter d. $Scale_Z$ is the pixel spacing in the Z-direction in $\mu m/pixel$. $Scale_Z$ is $3.86\mu m/pixel$ in the case of the Spectralis. These thickness profiles are then transformed back from the radial coordinate system of the interpolated B-Scan to the cartesian coordinates of the volume scan and color coded. As the Spectralis provides a hardware registered SLO-image besides the OCT-scan, the resulting 2D thickness map can be translucently laid over the SLO image to provide an instant RNFL thickness visualization. A color coding of the map enables the differentiation between the gray scale SLO image and the RNFL thickness values. Visual examples of the method are given in the results Section 3.

2.5. Evaluation

In order to quantitatively evaluate the algorithm we wrote a graphical user interface in Matlab (Mathworks, Inc., Natick, Massachusetts, USA) for displaying the segmentation results. Various display modes are available. For example, the image contrast can be adjusted and layer borders can be switched on and off. The simultaneously acquired SLO image of the Spectralis is also displayed. Manual corrections to the automated segmentation can be made by free-hand repainting of the segmentation borders. No region has to be selected. The method allows for the correction of even the smallest errors. We decided not to allow the observers to draw complete manual segmentations but rather to correct errors of the automatic segmentation. This was done for two reasons: Firstly, offsets in the segmentation lines often appear when they are completely manually drawn, even within one image and especially among different observers. This means that lines or line segments are shifted in the Z-direction by a few pixels distance. The human observers do not draw attention to this constant shift, but it prevents precise evaluation. The second reason for the correction of errors is that a complete manual segmentation of the dataset would be too time-consuming. The error correction alone took up to 12 hours for the entire circular B-Scan dataset for each reviewer.

Two experts in the field (authors MM and RT) and a student individually reviewed all segmented images of the circular B-Scan dataset. All automated segmentations were created with the same parameter set. A common rule on how to treat blood vessel regions was designed beforehand by the two experts. The segmentation boundary should follow the intensity drop at the ONFL as long as it is visible. Throughout the shadow region the two loose ends should be connected with a straight line. The student was trained on example images for the task.

To evaluate the automated segmentation three manual corrections were therefore available. To generate one segmentation per A-Scan that holds as a gold standard (GS) for the evaluation, the median border was selected from the three manual corrections. Thus, outlier corrections are not taken into account. The resulting gold standard is not an average, but lies exactly on a position where at least one observer set it. B-Scans consisting of less than 768 A-Scans were interpolated to 768 A-Scans before the evaluation computations. The differences between the segmentations were not calculated at the specific boundaries, but by looking at the resulting RNFL thickness at each position r. The RNFL thickness was calculated similar to Eq. (8).

For the evaluation of observer differences, an agreement between the observers is given. Agreement in our case defines the percentage of images were the mean absolute observer difference ($MAOD_i$) over all 768 A-Scans lies below a certain threshold. The $MAOD_i$ for a single

image with number i in the dataset is computed by:

$$MAOD_i = \frac{1}{768} \sum_{r=1}^{768} \frac{1}{3} (|RNFL_{Obs1,i}(r) - RNFL_{Obs2,i}(r)|$$
$$+ |RNFL_{Obs2,i}(r) - RNFL_{Obs3,i}(r)|$$
$$+ |RNFL_{Obs1,i}(r) - RNFL_{Obs3,i}(r)|) \tag{9}$$

$RNFL_{Obs1,i}(r)$ denotes the RNFL thickness profile of image number i computed out of the manually corrected ILM and ONFL (see Eq. 8) of observer 1. $RNFL_{Obs2,i}(r)$ and $RNFL_{Obs3,i}(r)$ are defined accordingly. MAOD without the subscript i denotes an average of the $MAOD_i$ over a part or the complete image data set. This convention holds also for similar B-Scan related measurements. The agreement is then:

$$Agreement = \frac{\#Img_{MAOD_i < t}}{\#Img} \tag{10}$$

The number of images is given by $\#Img$, the number of images with the $MAOD_i$ below a certain threshold t is given by $\#Img_{MAOD_i < t}$. If the averaging in the $MAOD_i$ formula is carried out over the image data set instead of the A-Scans of one image, we obtain the mean absolute observer difference for each A-Scan position ($MAOD(r)$):

$$MAOD(r) = \frac{1}{\#Img} \sum_{i=1}^{\#Img} \frac{1}{3} (|RNFL_{Obs1,i}(r) - RNFL_{Obs2,i}(r)|$$
$$+ |RNFL_{Obs2,i}(r) - RNFL_{Obs3,i}(r)|$$
$$+ |RNFL_{Obs1,i}(r) - RNFL_{Obs3,i}(r)|) \tag{11}$$

The difference of the automated segmentation to the gold standard ($Diff_i(r)$) for a single image with number i in the dataset is computed by:

$$Diff_i(r) = RNFL_{GS,i}(r) - RNFL_{autom,i}(r) \tag{12}$$

where $RNFL_{GS,i}(r)$ denotes the thickness profile according to the gold standard and $RNFL_{autom,i}(r)$ the thickness profile generated by the automated segmentation. We consider thickness profile differences below $8\mu m$ as negligible. If an A-Scan has less than $8\mu m$ thickness difference, we set it to $0\mu m$ for further evaluation. All subsequent numbers and graphs are calculated from these thresholded data. We use this threshold because it lies in the range of the resolution limit of the Spectralis in tissue. It roughly equals a 2 pixel distance on the image. No other exceptions are made, no image is excluded due to algorithm failure reasons.

The mean RNFL thickness($mRNFL_i$) is a parameter used for glaucoma diagnosis. It is computed by

$$mRNFL_i = \frac{1}{768} \sum_{r=1}^{768} RNFL_i(r) \tag{13}$$

The difference of the $mRNFL_i$ between the GS and the automatic segmentation (mean difference to the gold standard - MDG) is thus

$$MDG_i = \frac{1}{768} \sum_{r=1}^{768} Diff_i(r) \tag{14}$$

The averaging may cancel out positive and negative difference values. The mean *absolute* difference to the gold standard ($MADG_i$) is an additional error measurement and better represents
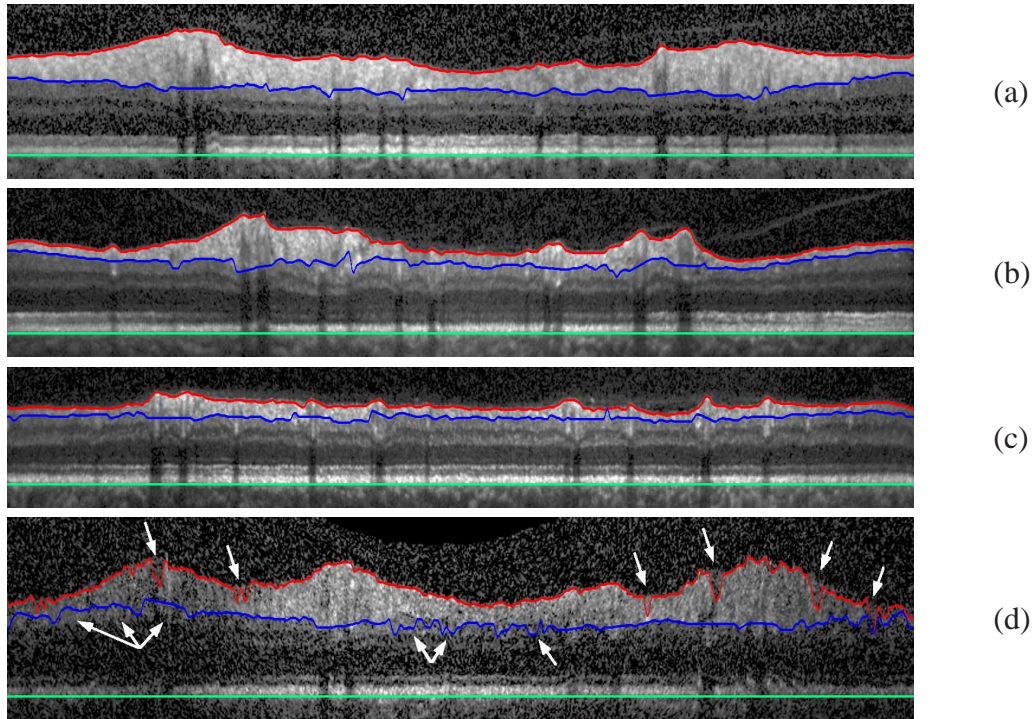
Fig. 6. Example results. (a) Normal eye. QI = 0.65. Automatically measured mean retinal nerve fiber layer thickness ($mRNFL_i$) = 111.47 μm (b) Glaucomatous eye. QI = 0.70. $mRNFL_i$ = 62.55 μm (c) Glaucomatous eye. QI = 0.67. $mRNFL_i$ = 42.16 μm (d) Normal eye. Very low image quality. QI = 0.54. $mRNFL_i$ = 111.90 μm. White arrows indicate segmentation errors.

the total quantitative segmentation error:

$$MADG_i = \frac{1}{768} \sum_{r=1}^{768} |Diff_i(r)|$$ (15)

This evaluation measurement can be, similar to the MAOD(r), computed for each A-Scan and over the entire data set instead of on a single image:

$$MADG(r) = \frac{1}{\#Img} \sum_{i=1}^{\#Img} |Diff_i(r)|$$ (16)

Replacing the $MAOD_i$ in Eq. (10) by the $MADG_i$ yields an agreement value between the GS and the automatic segmentation.

The evaluation results presented in Section 3 were generated after all algorithm parameters had been fixed. No parameter was changed during the evaluation procedure. All scans of normal subjects and glaucoma patients were processed with the same parameter set.

3. Results and Discussion

Fig. 6 shows visual examples of the segmentation results. Figure 6 (a) displays segmentation results of an image of a normal eye and (b) and (c) of glaucomatous eyes. The RNFL boundaries
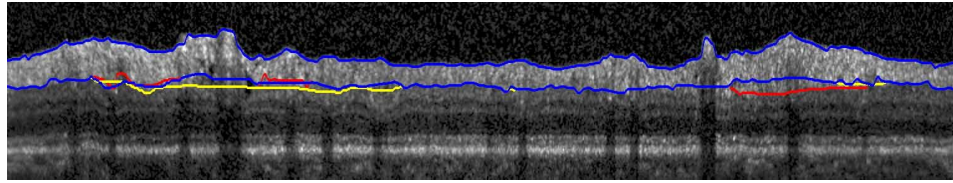
Fig. 7. Manually corrected segmentation from one reviewer (blue). The red and yellow lines indicate a differing correction from the other two reviewers.

are detected without severe errors in all these example cases, independent of the actual thickness of the RNFL.The significant variation of the RNFL thickness throughout the image is also captured. This is shown in Fig. 4 (e) and 6 (b). Figure 6 (d) shows a segmentation result of an image with very low quality. Here disruptions in the boundaries, as well as minor segmentation errors, can be observed and are marked with arrows. Note that, on the left and right side of the image clear boundaries can not be observed at all. The varying image quality that is best in the nasal quadrant is due to the signal fall-off of FD-OCT systems. The imaging range of an FD-OCT is limited by this signal fall-off. This is caused by the attenuation of the OCT signal due to the washout of the interferogram fringe visibility with increasing path-length difference of the interferometer [38]. The nasal quadrant of the example image in Fig. 6 (d) was placed on the upper side of the original image before aligning it to the RPE and thus has the shortest path-length difference.

The average runtime of the algorithm was measured on the complete circular B-Scan dataset using a MacBook Pro, Intel Core 2 Duo, 2,66 GHz with 4GB main memory. The code was written in Matlab (Mathworks, Inc.). Only one processor core was utilized. The average runtime was 20.5s. It did not differ substantially from normal subjects to glaucoma patients, or between images of good or bad quality. Included in the average runtime are the loading of the data from the hard disc and storing of the results. The biggest part of the running time (in average 73.5%) was used for the complex diffusion filter. Diffusion filters can be considerably optimized for speed by using multigrid technologies. Other parts of the algorithm can also be accelerated by implementing them in a more efficient language, for example C++. However, algorithm speed is not the focus of this paper. Therefore we did not optimize for computational efficiency.

Table 1. Agreement between the three manual corrections. Agreement in this work denotes the percentage of images where the mean absolute difference over all A-Scans lies below a certain threshold (see Section 2.5).

Threshold	Agr. 1-2 (%)	Agr. 1-3 (%)	Agr. 2-3 (%)	Average Agr.
5 μm	91.1	93.6	94.6	93.1
10 μm	97.5	99.0	98.5	98.4

We first quantitatively review the inter observer deviations. In Table 1 the agreements of the three observers are shown. There is a high degree of inter-observer agreement with an average 93.7% for a $5\mu m$ threshold and 98.2% for a $10\mu m$ treshold. On the other hand, these numbers show that the match is not perfect.

To localize the mismatch, in Fig. 8 the mean absolute observer difference for each A-Scan position ($MAOD(r)$) is plotted in addition with the blood vessel densitiy ($BVD(r)$) and the mean GS RNFL thickness ($mRNFL_{GS}(r)$). The $BVD(r)$ denotes the percentage of images, where a
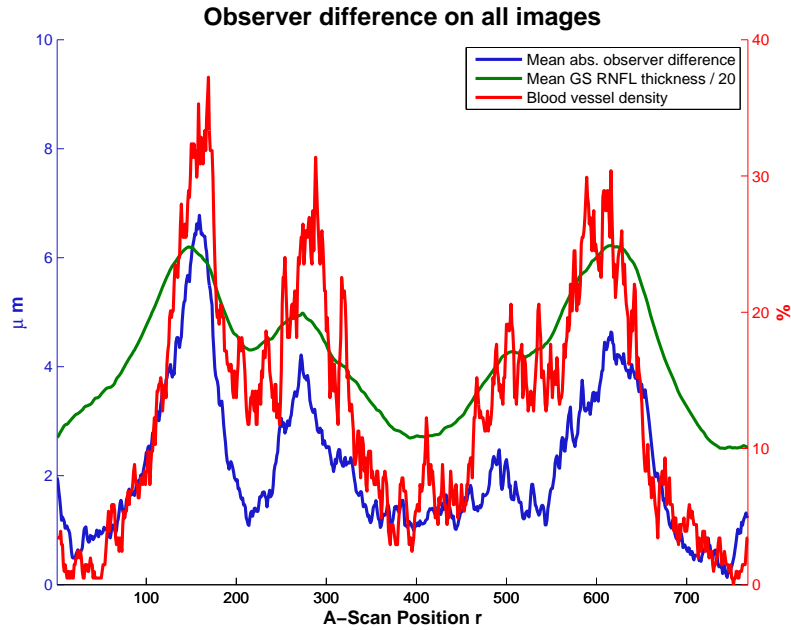
Fig. 8. Mean absolute observer difference ($MAOD(r)$: blue) at A-Scan position r. Mean gold standard retina thickness ($mRNFL_{GS}(r)$: green), scaled by a factor of $\frac{1}{20}$. Blood vessel distribution ($BVD(r)$: red). Herefore values are given on the right side and correspond to the percentage of images where a blood vessel is detected at the A-Scan position r. Correlation between $MAOD(r)$ and $BVD(r)$: 0.86, correlation between $BVD(r)$ and $mRNFL_{GS}(r)$: 0.87, correlation between $mRNFL_{GS}(r)$ and $MAOD(r)$: 0.84.

blood vessel is detected by the automatic algorithm at an A-Scan position r. All 3 curves are computed over the entire circular B-Scan dataset. It can be seen that the main peaks of all curves lie in the same positions. The correlation in between them is also high. Between $MAOD(r)$ and $BVD(r)$ it is 0.86, between the $BVD(r)$ and the $mRNFL_{GS}(r)$ 0.87 and between $mRNFL_{GS}(r)$ and $MAOD(r)$ 0.84. The observer differences are located mainly around blood vessel regions. This shows that, despite the fact that a rule was formulated for the treatment of blood vessel regions, there are still perceptual differences between the observers. The correlation of high RNFL thickness and high blood vessel density is due to the fact, that the OCT scans are taken on a circle of 3.4mm. Using this diameter, the nerve fibers and the main blood vessels are concentrated on the same positions. This does not necessarily hold for other scan circle diameters.

After localizing the inter-observer deviations, a second point of interest is whether normal and glaucomatous eyes were treated differently. The MAOD averaged over a) the 72 glaucoma patients and b) the 132 normal subjects separately is given in Table 2. It is higher for glaucoma patients (2.6μm) than for normal subjects (1.9μm). The difference becomes more evident when examining the mean difference measurements in relation to the mean RNFL thickness for each B-Scan. The relative mean absolute difference is more than twice as high for glaucoma patients (4.7%) compared to normal subjects (1.9%).

The influence of the image quality on the manual segmentations can also be viewed in Table 2. The 72 images with lowest quality (QI < 0.69) were selected from the database for this evaluation. This number was chosen so that it equals the number of images from glaucoma patients. The correlation of the images from glaucomtous eyes and low quality images can

Table 2. Number of images (#Img) in each group and the mean absolute observer differences (MAOD), averaged over all scans in the respective group ($\pm$ standard deviation). The MAOD is also shown with respect to the gold standard RNFL thickness. The numbers are calculated for the complete circular B-Scan dataset (All), the glaucoma patients and normal subjects, and the images of lowest quality (QI < 0.69) and high quality (QI $\geq$ 0.69).

Data	#Img	MAOD [μm]	rel. MAOD [%]
All	204	2.1 $\pm$ 2.0	2.9 $\pm$ 3.8
Glaucoma	72	2.6 $\pm$ 2.4	4.7 $\pm$ 5.7
Normal	132	1.9 $\pm$ 1.6	1.9 $\pm$ 1.5
Low QI	72	2.7 $\pm$ 2.2	3.5 $\pm$ 3.9
High QI	132	1.9 $\pm$ 1.8	2.6 $\pm$ 3.7

be computed in that way. The correlation is 0.12. We conclude that scans from both normal subjects and glaucoma patients were recorded with low quality. The MAOD on low quality images (2.7μm) was higher compared to that of better quality (1.9μm). This shows that, on a good quality scan, boundaries are placed more consistently among human observers.

Table 3. Number of images (#Img) in each group and the average evaluation results ($\pm$ standard deviation). The mean RNFL thickness (mRNFL), the mean difference to the gold standard (MDG), mean absolute difference to the GS (MADG) and the MADG in relation to the mean RNFL thickness computed out of the GS (given in %) are shown. The numbers are calculated for the complete circular B-Scan dataset (All), the glaucoma patients (Gl.) and normal subjects (Nor.), and the images of lowest quality (QI < 0.69) and high quality (QI $\geq$ 0.69).

Data	#Img	mRNFL [μm]	MDG [μm]	MADG [μm]	MADG [%]
All	204	84.0 $\pm$ 19.1	2.4 $\pm$ 3.4	3.5 $\pm$ 3.5	4.1 $\pm$ 5.5
Glaucoma	72	65.3 $\pm$ 15.7	0.9 $\pm$ 2.8	2.9 $\pm$ 3.5	4.9 $\pm$ 8.1
Normal	132	94.1 $\pm$ 11.7	3.2 $\pm$ 3.4	3.6 $\pm$ 3.4	3.7 $\pm$ 3.3
Low QI	72	83.6 $\pm$ 21.0	2.9 $\pm$ 4.3	4.5 $\pm$ 4.5	5.7 $\pm$ 8.3
High QI	132	84.1 $\pm$ 18.1	2.1 $\pm$ 2.7	2.7 $\pm$ 2.5	3.3 $\pm$ 2.8
Nor. & Low QI	40	97.6 $\pm$ 12.6	4.3 $\pm$ 4.3	4.9 $\pm$ 4.3	4.9 $\pm$ 4.2
Nor. & High QI	92	92.6 $\pm$ 11.0	2.7 $\pm$ 2.8	3.0 $\pm$ 2.8	3.1 $\pm$ 2.8
Gl. & Low QI	32	66.1 $\pm$ 15.4	1.1 $\pm$ 3.7	3.9 $\pm$ 4.8	6.7 $\pm$ 11.6
Gl. & High QI	40	64.6 $\pm$ 16.1	0.7 $\pm$ 1.9	2.1 $\pm$ 1.4	3.5 $\pm$ 2.8

We also compared the automated segmentations to the gold standard. We first compare the mean RNFL thickness of the GS and the automatic segmentation, as this value serves as a glaucoma parameter in daily cinical routine. The average difference to the gold standard (MDG) over the whole circular B-Scan dataset is 2.4 μm (see Table 3). On glaucoma patients it is 0.9 μm compared to 3.2 μm on healthy eyes. On good quality images it is 2.1 μm compared to 2.9 μm on the low quality images. Note that the average error in all groups is positive. One can conclude that the error is systematic. If errors occur in the segmentation, there is a high probability that the RNFL thickness is on average segmented too thinly.

Despite that, as stated in Section 2.5, averaged positive and negative errors at different A-Scan positions may cancel each other out at least partially when using the MDG as an error

Table 4. Agreement between the gold standard (per A-Scan median of the 3 reviewers) and the automated segmentation. Agreement in this work denominates the percentage of images where the mean absolute difference per A-Scan lies below a certain threshold (see 2.5). The agreement is calculated for the complete circular B-Scan dataset (All), the glaucoma patients (72 images) and normal subjects (169 images) and the images of lowest quality (72 images, QI < 0.69) and high quality (169 images, QI ≥ 0.69).

Data	5 μm Th. [%]	10 μm Th. [%]
All	82.3	95.1
Glaucoma	90.0	97.2
Normal	78.0	94.0
Low QI	75.0	91.7
High QI	86.3	97.0
Normal & Low QI	67.5	90.0
Normal & High QI	82.6	95.6
Glaucoma & Low QI	83.9	93.5
Glaucoma & High QI	95.1	100

measurement. Hence Table 3 also shows the mean *absolute* differences to the gold standard averaged over all A-scans (MADG). Glaucoma patients have smaller MADG ($2.9\mu m$) than normal subjects ($3.6\mu m$). If these values again are examined in relation to the mean RNFL thickness, the behavior changes. The average relative error on glaucoma patients is higher (4.9%) than on normal subjects (3.7%). On images with low quality both the absolute and the relative error ($4.5\mu m - 5.7\%$) are higher than on images of good quality ($2.7\mu m - 3.3\%$). This shows that the algorithm results are affected by the presence of pathologies and by low image quality. The effect adds up at the intersection of these two groups (glaucoma and low quality images) with an error of $3.9\mu m - 6.7\%$. These numbers, however, also state that the effect is not severe and lies within the standard deviations of the measurements in the respective groups. The numbers for all other possible intersection groups are given in Table 3.

The agreement of the automated segmentation with the gold standard shown in Table 4 are generally lower than the inter-observer agreements. If a threshold of $10\mu m$ for the MADG is set, 95% of the images show an agreement. Suprisingly, the agreement in the glaucoma group is higher than on scans of normal subjects. 97.2% of the glaucomatous images have an average absolute error of at most $10\mu m$. As it can be concluded already out of the difference measurements, low image quality results in lowered agreement.

A point of special interest is the location of the differences. Figure 9 shows (similar to Fig. 8) the mean absolute difference to the gold standard at A-Scan position ($MADG(r)$) computed over the whole circular B-Scan dataset. The blood vessel distribution is also shown. They correlate with a factor of 0.84. Most differences are found in the superior quadrant and the inferior quadrant (quadrant positions see Fig. 1). This behavior showed up in the inter observer review (see Fig. 8), too. The segmentation errors concentrate in regions with a high blood vessel density. The treatment of blood vessels in automated segmentations is still an open question, though (see Hood et al. [39]). It can lead to segmentation differences among OCT system implementations (see Hood et al. [40]). Our proposed segmentation algorithm generates a thinner RNFL thickness than the current software of the Spectralis in normal subjects. Baleanu et al. [35] reported a mean RNFL thickness of $97.2 \pm 9.7\mu m$ for healthy eyes, while our algorithm measured $94.1 \pm 11.7\mu m$ (see Table 3).

To conclude the evaluation of the automated segmentation on circular B-Scans, a compar-
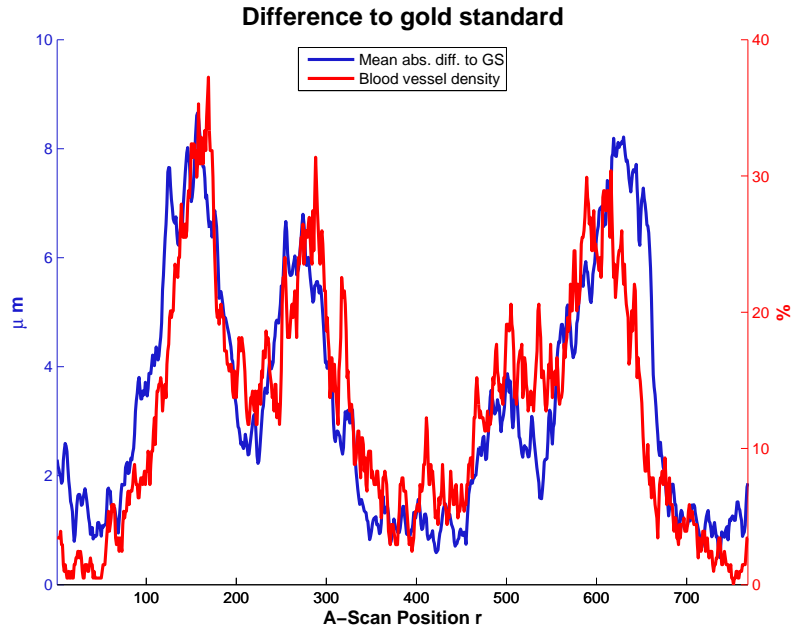
Fig. 9. Mean absolute difference to gold standard per A-Scan ($MADG(r)$: blue) computed over the whole circular B-Scan dataset. Blood vessel distribution ($BVD(r)$: red). Values herefore are given on the right side and correspond to the percentage of images where a blood vessel is detected at the certain A-Scan position r. Correlation between $MADG(r)$ and $BVD(r)$: 0.84

ison of the $MADG(r)$ to the $MAOD(r)$ is performed. As Fig. 10 shows, the difference of the automated segmentation to the gold standard is higher than the inter-observer difference. But one can also notice that it scales almost linearly. This is an indicator that the automated segmentation fails mostly in regions where a definite objective decision by humans is also hard to obtain. The correlation between the $MADG(r)$ and $MOAD(r)$ calculated over the whole circular B-Scan dataset is 0.93.

The usability of the proposed segmentation algorithm for 3D Volume scans is shown exemplarily on two volume scans. As described in Section 2.4 circular B-Scans with varying diameters were interpolated out of the volumes. As the data density in the Y-direction of the volume is coarse, compared to the X- and Z-direction, and motion compensation is carried out by the Spectralis, a lower image quality is expected. This assumption is supported by a visual inspection. In Fig. 11 an interpolated scan is shown compared to a circular B-scan from a glaucoma patient. Both the interpolated and the circular scan have the same diameter. The interpolated scan suffers from interpolation artifacts especially in the nasal and temporal quadrant. The data density in these regions is the lowest (see arrows in Fig. 11 (b)). When looking at the transition between the inferior and nasal quadrant it must be noted that some image features are shifted. This shift is inherent in the data and not produced by the interpolation algorithm.

Despite the lower quality of the interpolated image and the feature shifts, the RNFL thickness plots of the automated segmentations show a high correlation of 0.82 and low difference in mean RNFL thickness of 2.9 μm. As the algorithm does not make assumptions on the RNFL thickness, it is applied to the interpolated scans without changing any parameter. If now multiple
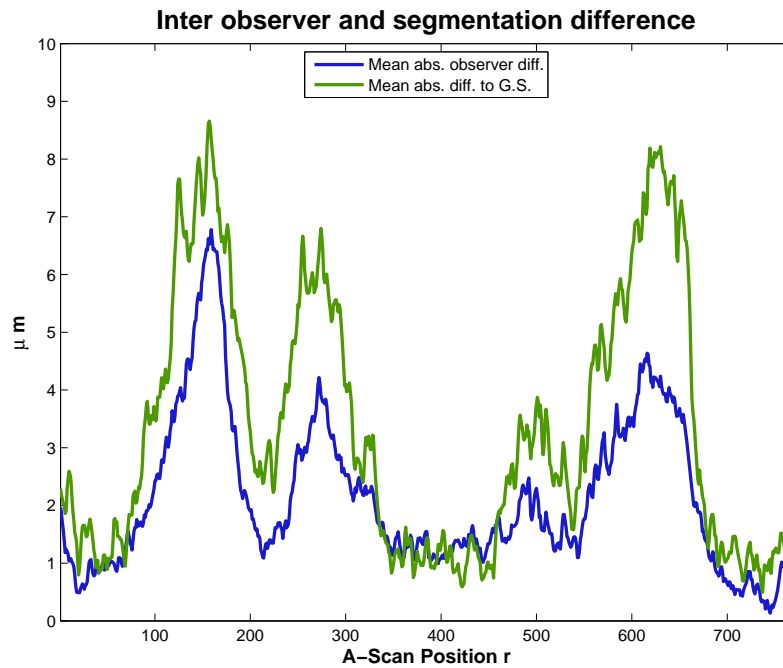
Inter observer and segmentation difference

Fig. 10. Mean absolute observer difference ($MAOD(r)$: blue) compared to mean absolute difference to gold standard per A-Scan ($MADG(r)$: green) over the entire circular B-Scan data set. The correlation between the plots is 0.93
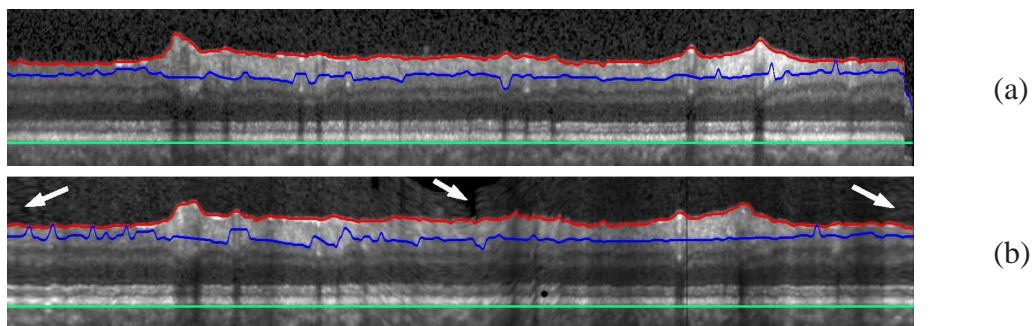

(a)

(b)

Fig. 11. (a) Circular B-Scan of a glaucoma patient. (b) Circular B-Scan interpolated out of a volume scan of the same patient. Automated segmentation results are shown. The white arrows indicate regions with low image quality due to lower data density in the interpolation. The corresponding RNFL thickness plots are shown in Fig. 12.
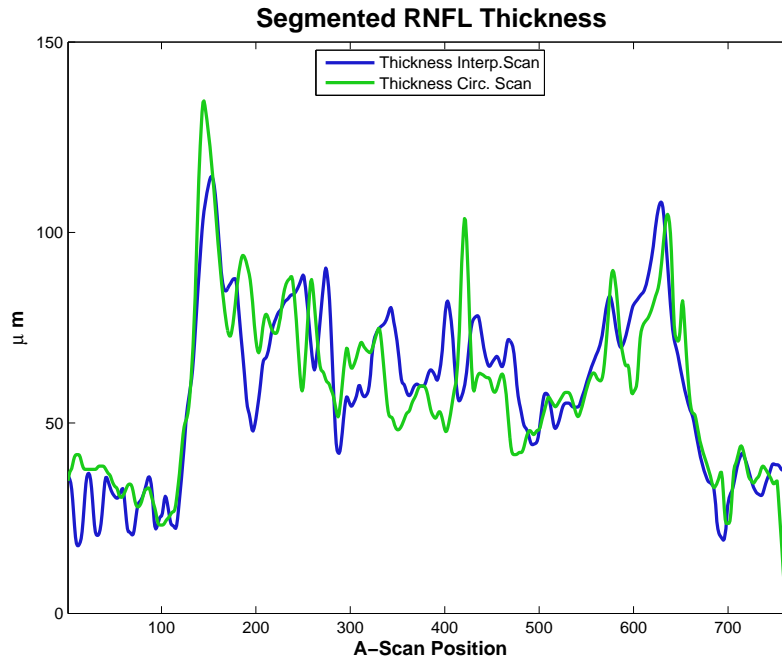
Segmented RNFL Thickness

Fig. 12. Retinal nerve fiber layer thickness (RNFL) from the automated segmentation. Circular B-Scan of a glaucoma patient: green, Mean thickness = 58.0 μm. Circular B-Scan interpolated out of a volume scan of the same patient: blue, Mean thickness = 55.1 μm. The correlation between both plots is 0.82, the thickness difference of the means is 2.9 μm.

interpolated scans with varying diameter are segmented and the resulting RNFL thickness is color coded and mapped back to the coordinate system of the volume scan, a thickness map is obtained. This map can be translucently laid over the SLO image acquired during the volume scan protocol of the Spectralis. An example thickness map of the same glaucoma patient as in Fig. 11 is shown in Fig. 13 (b). In Fig. 13 (a) a scan of a normal eye is shown in comparison. 100 circular scans with radii between 1 and $3mm$ are interpolated out of the volumes for these visualizations. This visualization shows the local RNFL loss of the glaucoma patient in the temporal quadrant. A visualization artifact is present in both scans in the nasal region due to the boundary segmentation differences in the B-scan. This artifact does not affect the instantaneous qualitative impression an observer can get from the thickness map.

4. Summary and Conclusion

We presented an automated algorithm for segmenting the RNFL on circular OCT B-Scans. Its key idea is the minimization of an energy term that takes into account the gradient along an A-Scan, as well as local and regional smoothness. The initialization segmentation provided to the minimization procedure is found by denoising the image with complex diffusion and using a heuristic selection of prominent edges. The algorithm makes only few general assumptions on the position and shape of the layer borders and thus can be applied on scans of both normal and glaucomatous eyes.

We have substantiated the applicability of the proposed algorithm by an evaluation on a dataset of images from 132 normal subjects and 72 glaucoma patients. The images were com-
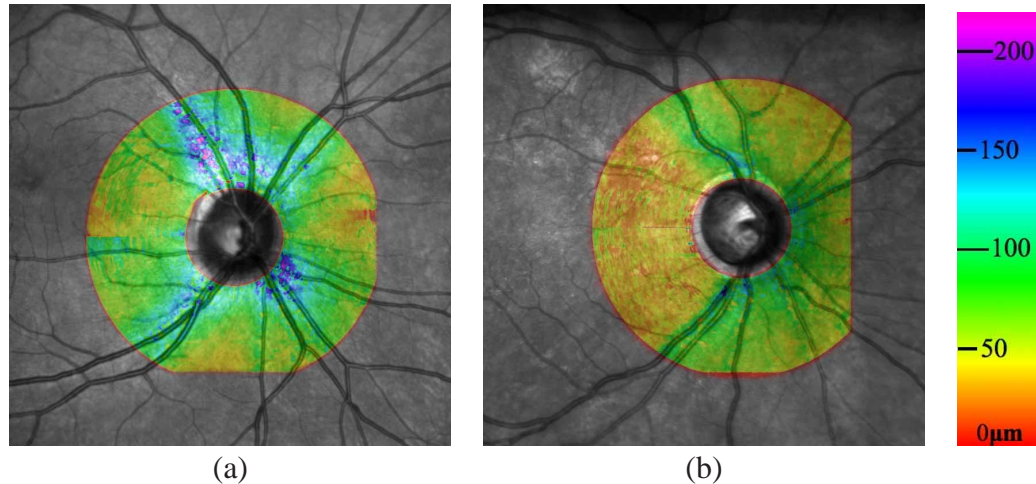
Fig. 13. Thickness maps overlaid with SLO image. (a) Normal subject. (b) Glaucoma patient with local RNFL loss. The RNFL thickness is visualized by a pseudo color scale, ranging from blue ($220\mu m$) over green to red ($0\mu m$). The used color scale is shown on the right.

pared to a gold standard generated out of manual corrections to the automated segmentations of two experts and a trained student. The mean absolute difference to the gold standard ($MADG_i$) lies below $10\mu m$ in 97.2% of all glaucomatous eyes and 94.0% of all normal eyes. 97.0% of the images with high quality have an overall $MADG_i$ below $10\mu m$ compared to 91.7% of images with low quality. The inter-observer difference is lower than the one between the automated segmentation and the gold standard, but scales nearly lineary (correlation coefficient 0.93). This leads to the conclusion, that the algorithm produces errors in positions where human observers also do not perfectly agree on. By interpolating circular scans out of OCT volumes the 2D algorithm is also applicable on 3D data. Using multiple scans with varying diameters, a 2D RNFL map can be generated. Relative measurement errors are slightly higher for pathologic data and low image quality. They are mostly located in regions with high blood vessel concentration. Thus, a more accurate segmentation at the position of the retinal vessels will increase the performance of the algorithm. However, the average absolute measurement error, as well as the average relative error, compared to the mean RNFL thickness, are very low. They are below the standard deviation measured in the glaucoma and normal group separately.

We believe that the automated segmentation algorithm for FD-OCT data presented and evaluated in this work provides a reliable tool for extracting diagnostically relevant parameters from circular B-Scans. The same algorithm applied to 3D volume data gives the physician an immediate impression of the RNFL thickness around the optic nerve head.

5. Acknowledgement

Table 5. Table of abbreviations (top) and symbols (bottom) in alphabetical order.

Abbreviation/Symbol	Explanation
BVD	Blood Vessel Distribution
BVR	Region in between two blood vessels
FD	Frequency Domain
I	Inferior
ILM	Inner limiting membrane
INL	Inner Nuclear Layer
IPL	Inner Plexiform Layer
ISG	Inner segment of the retina
MADG	Mean Absolute Differece to the Gold Standard
MAOD	Mean Absolute Observer Difference
MDG	Mean Difference to Gold Standard
mRNFL	Mean Retinal Nerve Fiber Layer Thickness
N	Nasal
OCT	Optical Coherence Tomography
ONFL	Outer Nerve Fiber Layer Boundary
ONL	Outer Nuclear Layer
OPL	Outer Plexiform Layer
OSG	Outer segment of the retina
QI	Quality Index
PS	Part Sum Ratio
RNFL	Retinal Nerve Fiber Layer
RPE	Outer Retinal Pigment Epithelium Boundary
Spectralis	Spectralis HRA+OCT (Heidelberg Engineering, Heidelberg, Germany)
S	Superior
SLO	Scanning Laser Ophthalmoscope
SVM	Support Vector Machine
TD	Time Domain
T	Temporal
VOL	Raw data format of the Spectralis HRA+OCT
$D(r)$	Regional smoothness term at A-Scan (r)
$E(r)$	Energy function at A-Scan (r)
$G(z,r)$	Gradient at image position (z,r)
$I(z,r)$	Image intensity at image position (z,r)
$MADG(r)$	Mean Absolute Differece to the Gold Standard at A-Scan (r)
$MAOD(r)$	Mean Absolute Observer Difference at A-Scan (r)
$N(r)$	Local smoothness term at A-Scan (r)
$Scale_Z$	Pixel Spacing in Z-direction in $\mu m/pixel$
$RNFL(r)$	Retinal Nerve Fiber Layer Thickness Profile at A-Scan r
α	Weighting factor
β	Weighting factor
σ	Standard deviation
σ_{CD}	Noise standard deviation estimate for complex diffusion
$\#N$	Number of pixels in a scan

Table 6. Overview (first part) over published research in the field of retina and retinal nerve
fiber layer (RNFL) segmentation on OCT data. Abbreviations see caption table 7.

Author	Year	Objective	Method	Data	Evaluation
Koozekanani et al. [8]	2001	Retina seg.	Edge detection. Regularization by a Markov model	1450 TD-OCT B-Scans scans from normal eyes	Quantitative evaluation with manually corrected segmentations
Ishikawa et al. [10]	2002	RNFL seg.	Edge detection with integritiy check	TD-OCT circular B-Scans: 86 scans from 21 NS, 131 scans from 32 OHP, 184 scans from 45 GP	Quantitative evaluation by marking errors
Fernandez et al. [14]	2005	7 layers seg.	Complex diffusion and coherence enhanced diffusion followed by edge detection	TD-OCT B-Scans: 72 scans from NS, scans of 4 different pathologic cases	Visual inspection
Ishikava et al. [11]	2005	5 layers seg.	Edge detection with integritiy check	TD-OCT circular B-Scans: 144 scans from 24 NS, 144 from 24 GS included.	Quantitative evaluation by marking errors. Exclusion of bad quality images.
Mujat et al. [16]	2005	RNFL seg.	Anisotropic noise suppression and deformable splines	SD-OCT volumes of NS	Visual inspection
Shahidi et al. [13]	2005	3 layer groups seg.	Averaging A-Scans and egde detection	TD-OCT B-Scans of 10 NS	Reproducibility
Haecker et al. [21]	2006	ILM, RPE seg.	3D geometric graph cut and a priori contraints	TD-OCT radial scan sets: 9 scan sets from NS, 9 from PP	Qualitative evaluation by marking errors
Baroni et al. [18]	2007	2 layer groups seg.	Maximization of a likelihood function consisting of a gradient and local smoothness term	TD-OCT B-Scans: Scans of 18 NS, scans of 16 CCMP	Parameter adaption and error judging by 2 reviewers
Fuller et al. [31]	2007	Multiple or single layer seg.	SVM classifier training for each volume out of manually drawn regions	SD-OCT volumes of NS and patients	Segmentation time evaluation. Comparison to manual seg.
Joeres et al. [33]	2007	Retina, OPL and subretinal tissue seg.	Manual seg. with OCTOR software	TD-OCT B-Scans of 60 AMD patients	Repeatablity and agreement of two operators
Sadda et al. [34]	2007	Retina seg.	Manual seg. with OCTOR software	TD-OCT B-Scans of patients with macular diseases	Repeatablity and agreement of two operators
Somfai et al. [17]	2007	Effect of operator error on seg.	Analysis with custom [14] and commercial software	TD-OCT B-Scans of 8 NS and 1 DME patient. 4 scans with different operator errors per person.	Comparison of optimal automatic seg. with seg. on images with worse quality
Szulmowski et al. [32]	2007	Group of posteriour layers seg.	Classifier training out of manually drawn regions	SD-OCT volume data of NS and patients	Visual inspection

Table 7. Overview (second part) over published research in the field of retina and retinal nerve fiber layer (RNFL) segmentation on OCT data. Abbreviations: Segmentation (seg.), normal subject (NS), ocular hypertension patient (OHP), glaucoma patient (GP), papilledema patient (PP), outer photoreceptor layer (OPL), age related macula degenration (AMD), diabetic macula edema (DME), optic neuropathy patient (ONP), perimetric glaucoma patient (PGP), preperimetric glaucoma patient (PPGP). The table does not claim to be complete.

Author	Year	Objective	Method	Data	Evaluation
Garvin et al. [22]	2008	5 layers seg.	3D geometric graph cut and a priori contraints	TD-OCT radial scan sets from 12 ONP. 1 diseased eye and 1 normal from each patient	Qualitative evaluation using manual seg. by 3 observers
Götzinger et al. [30]	2008	RPE seg.	Two algorithms with different complexity	SD-PS-OCT volumes of NS and patients	Visual inspection
Tolliver et al. [24]	2008	RNFL, RPE seg.	Boundary detection by spectral rounding	SD-OCT volumes of 2 NS and 9 patients	Quantitative evaluation using manual seg.
Tan et al. [19]	2008	5 layers seg.	Progressive edge detection, each step less A-Scan averaging	TD-OCT B-Scans of 44 NS, 73 PGP and 29 PPGP	Exclusion of scans with seg. errors in the study
Mishra et al. [25]	2009	All (10) intraretinal layer seg.	Approximation and refinment of layer positions with dynamic programming	SD-OCT B-Scans of healthy and diseased rat retinas	Visual inspection
Tan et al. [20]	2009	2 layer groups seg.	Edge detection with 3D neighbor constraints and knowledge model	SD-OCT volume scans of 65 NS, 78 PGP and 52 PPGG	Exclusion of scans with seg. errors in the study
Yazdanpanah et al. [26]	2009	5 layers seg.	Active contours: Minimization of an energy functional with a shape prior	20 SD-OCT B-Scans of rat eyes	Quantitative evaluation with manual segmentation
Chiu et al. [28]	2010	7 layers seg.	Graph theory and dynamic programming	SD-OCT Scans of 10 NS	Quantitative evaluation with manual seg. by 2 observers
Kajić et al. [27]	2010	9 layers seg.	Model based segmentation with shape and texture features	SD-OCT volumes of 17 normal eyes	Quantitative evaluation with manual seg. by 2 observers
Quellec et al. [23]	2010	10 layers seg., abnormality detection	Seg. see [22], texture and thickness features for abnormality detection	SD-OCT volumes of 13 NS	Quantitative evaluation with manual seg. by 2 observers
Vermeer et al. [29]	2010	5 layers seg.	Pixelwise classification with SVM, Level Set regularization	SD-OCT volumes of 10 NS and 8 GP	Quantitative evaluation with manual seg. of 1-2 B-Scans per volume