Research Article | Vol. 8, No. 3 | 1 Mar 2017 | BIOMEDICAL OPTICS EXPRESS 1874

Biomedical Optics EXPRESS

Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context

ALESSIO MONTUORO,[1,2,*] SEBASTIAN M. WALDSTEIN,[1,2] BIANCA S. GERENDAS,[1,2] URSULA SCHMIDT-ERFURTH,[1,2] AND HRVOJE BOGUNOVIĆ[2]

[1] *Vienna Reading Center, Department of Ophthalmology and Optometry, Medical University of Vienna, Austria*
[2] *Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology, Medical University of Vienna, Austria*
[*] *alessio.montuoro@meduniwien.ac.at*

Abstract: Modern optical coherence tomography (OCT) devices used in ophthalmology acquire steadily increasing amounts of imaging data. Thus, reliable automated quantitative analysis of OCT images is considered to be of utmost importance. Current automated retinal OCT layer segmentation methods work reliably on healthy or mildly diseased retinas, but struggle with the complex interaction of the layers with fluid accumulations in macular edema. In this work, we present a fully automated 3D method which is able to segment all the retinal layers and fluid-filled regions simultaneously, exploiting their mutual interaction to improve the overall segmentation results. The machine learning based method combines unsupervised feature representation and heterogeneous spatial context with a graph-theoretic surface segmentation. The method was extensively evaluated on manual annotations of 20,000 OCT B-scans from 100 scans of patients and on a publicly available data set consisting of 110 annotated B-scans from 10 patients, all with severe macular edema, yielding an overall mean Dice coefficient of 0.76 and 0.78, respectively.

OCIS codes: (100.6890) Three-dimensional image processing; (100.0100) Image processing; (170.4470) Ophthalmology; (170.4500) Optical coherence tomography.

References and links

1. R. D. Jager, W. F. Mieler, and J. W. Miller, "Age-related macular degeneration," New Engl. J. Med. **358**, 2606–2617 (2008).
2. J. A. Davidson, T. A. Ciulla, J. B. McGill, K. A. Kles, and P. W. Anderson, "How the diabetic eye loses vision," Endocrine **32**, 107–116 (2007).
3. P. A. Campochiaro, L. P. Aiello, and P. J. Rosenfeld, "Anti-vascular endothelial growth factor agents in the treatment of retinal disease: From bench to bedside," Ophthalmology **123**, S78–S88 (2016).
4. K. A. Rezaei and T. W. Stone, "2016 Global Trends in Retina Survey," American Society of Retina Specialists, Chicago, IL. (2016).
5. U. Schmidt-Erfurth, S. M. Waldstein, G.-G. Deak, M. Kundi, and C. Simader, "Pigment epithelial detachment followed by retinal cystoid degeneration leads to vision loss in treatment of neovascular age-related macular degeneration," Ophthalmology **122**, 822–832 (2015).
6. S. M. Waldstein, J. Wright, J. Warburton, P. Margaron, C. Simader, and U. Schmidt-Erfurth, "Predictive value of retinal morphology for visual acuity outcomes of different ranibizumab treatment regimens for neovascular AMD," Ophthalmology **123**, 60–69 (2016).
7. C. K. Hitzenberger, P. Trost, P.-W. Lo, and Q. Zhou, "Three-dimensional imaging of the human retina by high-speed optical coherence tomography," Opt. Express **11**, 2753–2761 (2003).
8. A. Salinas-Alamán, A. García-Layana, M. J. Maldonado, C. Sainz-Gómez, and A. Alvárez-Vidal, "Using optical coherence tomography to monitor photodynamic therapy in age related macular degeneration," Am. J. Ophthalmol. **140**, 23.e1 (2005).
9. S. M. Waldstein, A.-M. Philip, R. Leitner, C. Simader, G. Langs, B. S. Gerendas, and U. Schmidt-Erfurth, "Correlation of 3-dimensionally quantified intraretinal and subretinal fluid with visual acuity in neovascular age-related macular degeneration," JAMA Ophthalmol. **134**, 182–190 (2016).
10. 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**, 19413–19428 (2010).

#279865 https://doi.org/10.1364/BOE.8.001874
Journal © 2017 Received 2 Nov 2016; revised 1 Feb 2017; accepted 10 Feb 2017; published 27 Feb 2017

Biomedical Optics EXPRESS

11. A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool", BMC Med Imaging **15**, 29 (2015).
12. S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," Biomed. Opt. Express **6**, 1172–1194 (2015).
13. X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abramoff, and M. Sonka, "Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: Probability constrained graph-search-graph-cut," IEEE Trans. Med. Imaging **31**, 1521–1531 (2012).
14. M. K. Garvin, M. D. Abramoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images," IEEE Trans. Med. Imaging **28**, 1436–1447 (2009).
15. T. Schlegl, S. M. Waldstein, W.-D. Vogl, U. Schmidt-Erfurth, and G. Langs, "Predicting semantic descriptions from medical images with convolutional neural networks" Inf. Process. Med. Imaging **24**, 437–448 (2015).
16. Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. de Sisternes, and D. L. Rubin, "Automated drusen segmentation and quantification in SD-OCT images," Med. Image Anal. **17**, 1058–1072 (2013).
17. I. Oguz, L. Zhang, M. D. Abrámoff, and M. Sonka, "Optimal retinal cyst segmentation from OCT images," in "Medical Imaging 2016: Image Processing," M. A. Styner and Elsa D. Angelini, eds., Proc. SPIE **9784**, 97841E (2016)
18. J. Wang, M. Zhang, A. D. Pechauer, L. Liu, T. S. Hwang, D. J. Wilson, D. Li, and Y. Jia, "Automated volumetric segmentation of retinal fluid on optical coherence tomography," Biomed. Opt. Express **7**, 1577-1589 (2016).
19. S. P. K. Karri, D. Chakraborthi, and J. Chatterjee, "Learning layer-specific edges for segmenting retinal layers with large deformations," Biomed. Opt. Express **7**, 2888–28901 (2016).
20. M. Zhang, J. Wang, A. D. Pechauer, T. S. Hwang, S. S. Gao, L. Liu, Li Liu, S. T. Bailey, D. J. Wilson, D. Huang, and Y. Jia, "Advanced image processing for optical coherence tomographic angiography of macular diseases," Biomed. Opt. Express **6**, 4661–4675 (2015).
21. Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," IEEE Trans. Pattern Anal. Mach. Intell. **32**, 1744–1757 (2010).
22. L. Breiman, "Random forests," Mach. Learn. **45**, 5–32 (2001).
23. A. Montuoro, C. Simader, G. Langs, and U. Schmidt-Erfurth, "Rotation invariant eigenvessels and auto-context for retinal vessel detection," in "Medical Imaging 2015: Image Processing," S. Ourselin and M. A. Styner, eds., Proc. SPIE **9413** 94131F (2015)
24. D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," Int. J. Comput. Vision **77**, 125–141 (2008).

1. Introduction

Macular edema is a swelling or thickening of the macula, the area of the retina responsible for central vision. It occurs frequently as secondary to age-related macular degeneration (AMD), diabetic retinopathy (DR) or retinal vein occlusion (RVO). AMD alone is the leading cause of irreversible blindness in people over 50 years in the developed world [1] while diabetic macular edema (DME) is one of the leading causes of blindness in the United States [2]. The swelling is mainly the result of the accumulation of fluid inside (intraretinal fluid - IRF) and underneath the neurosensory retina (subretinal fluid - SRF), which severely affects the otherwise well defined layered structure of the retina and can lead to profound loss of vision. An example of an optical coherence tomography (OCT) slice of a retina with macular edema is shown in Fig. 1, with IRF depicted in white and SRF depicted in blue.

Intravitreal anti-vascular endothelial growth factor (anti-VEGF) therapy is an effective and safe treatment option for patients suffering from these conditions [3]. Most physicians employ an individualized therapeutic regimen, which aims at treating as little as possible to avoid associated morbidity and cost, but as much as needed to control the chronic disease [4]. Recent studies have shown that individual retinal morphology has predictive value for treatment requirements and prognosis, which could lead the way to personalized treatment regimes, reducing the burden on patients and the health care system [5, 6]. This highlights the need for robust and sensitive quantitative imaging biomarkers, on which treatment decisions could be reliably based.

Spectral domain OCT [7] is a powerful and widely used modality for 3D imaging of the retina in vivo with μm resolution. Due to its capability to visualize the retina and its layers with fluid pockets in fine detail, it plays a vital role in clinical decision making as well as in large scale

Biomedical Optics EXPRESS

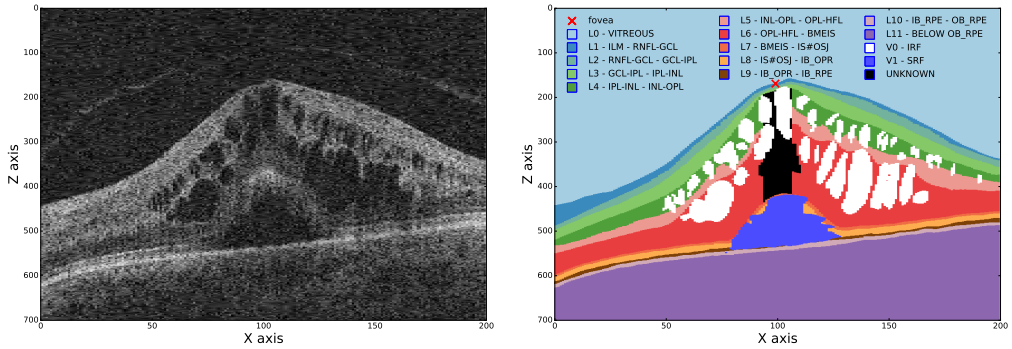

Fig. 1. **Left:** B-scan of a SD-OCT volume of a patient with pronounced macular edema. Note the loss of OCT signal below highly absorbing regions such as fluid. **Right:** Voxel-wise manual annotation of 14 regions.

clinical trials. Nevertheless, the ever growing amount of image acquisitions together with the steadily increasing spatial resolution of OCT scans produce an amount of data that makes their manual analysis prohibitively time-consuming. This creates an unmet need for the development of fully automated image segmentation methods to quantify the status of retinal layers and fluid in a streamlined, objective and repeatable way.

An imaging biomarker that has already been proven to be linked with retinal function and treatment response is the thickness of individual retinal layers [8]. Likewise, the quantification of IRF and SRF volumes has recently been shown to be clinically relevant [9]. However, while the retinal layers are easily discernible in healthy and moderately diseased retinas, the presence of IRF and SRF disrupts their visibility. In addition, highly absorbing materials in the retina (such as hemorrhage and lipid exudation) result in further degradation of OCT image quality below such lesions (Fig. 1), making the retinal layers barely distinguishable. Thus, the segmentation of patients with macular edema remains a very challenging task both for expert readers and particularly for automated methods.

In terms of related work, both retinal layer segmentation algorithms and fluid segmentation algorithms have been previously proposed. However, up until recently these algorithms segmented the imaging biomarkers either individually or consecutively (e.g. the retinal layers first and then IRF in the volume bounded by certain retinal layers). Thus, the simultaneous segmentation of fluid and layers could exploit their complex interaction and yield improved segmentation results on severely diseased cases. Nevertheless, the high degree of variability in the appearance of such cases makes the accurate modeling of this fluid-layer interaction challenging. The layer segmentation approach proposed in [10] has recently been extended to additionally perform fluid segmentation [12] in an iterative fashion. However it is performed in 2D, does not differentiate between the fluid types, and it is limited to eight main layers only. Another notable method is the graph-theoretic approach proposed in [13] that combines a graph-search layer segmentation approach [14] with graph-cut fluid segmentation. However, it segments only the inner and outer boundaries of the retina and does not differentiate between the fluid types. In a similar fashion, a number of fluid segmentation approaches have been proposed that either require or perform a prior layer segmentation in order to improve the overall segmentation accuracy [15–18]. The dynamic programming based approach presented in [19] uses a machine learning algorithm to predict the cost function for the final segmentation and is evaluated on the same data set as used in [12]. However, the method does not segment fluid, hence for the validation the layers inside the fluid areas were created using a linear interpolation, making it hard to compare with methods that segment both fluid and layers. The method proposed in [20] uses a directional

Biomedical Optics EXPRESS

graph search with a manually constructed cost function for the retinal layer segmentation and employs a semi-automatic segmentation in cases with large morphologic deformations.

The aim of this work is to develop a fully automated segmentation of retinal structures in the presence of macular edema using a data-driven approach by learning image and spatial features from a training set of manually labeled OCT images. In contrast to the related work, in this paper we propose a method that performs a simultaneous 3D segmentation of all eleven retinal layers together with the two fluid regions (IRF and SRF). After an initial voxel classification step the retinal layers are segmented using a graph-theoretic approach [14], and the results are then iteratively refined by exploiting the complex interaction between different retinal structures using auto-context methodology [21].

2. Method

2.1. Definitions

OCT acquires a single axial scan (A-scan, Z axis in Fig. 2) at a time and reconstructs a series of cross sectional slices (B-scans, X/Z axis) by scanning through the volume. We define a surface S to be a terrain like, continuous boundary that splits the OCT volume into two parts, one above the surface (called foreground F) and one below the surface (called background B). We further define a layer (L) to be the volume enclosed by two surfaces (e.g. $L5 = [S4, S5]$), and two fluid filled volumes ($V0$ and $V1$ corresponding to IRF and SRF, respectively). The definition of the used surfaces, regions and their structural relationship is determined by the anatomy of the retina and is provided in Table 1, resulting in 14 regions in total.

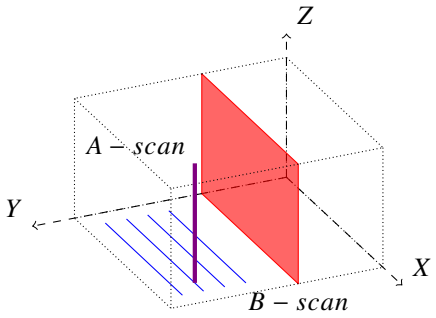

Fig. 2. OCT acquisition and the coordinate system. 1D axial scans (A-scan, purple) are combined to form a 2D cross sectional slice (B-scan, red) by scanning through the volume in a raster scan pattern (blue). Multiple B-scans are then combined to form a complete OCT volume.

2.2. Workflow overview

The overall workflow (Fig. 3) of the proposed method consists of the following steps: First, image-based features are extracted from the raw OCT data and are used together with manual labels to train a base voxel classifier. The resulting probability map is then used to perform the initial surface segmentation and to extract various context-based features. Second, these features are used in conjunction with the image-based features to train another classifier. Such context-based feature extraction and additional classifier training is then iteratively repeated multiple times in auto-context loop.

Biomedical Optics EXPRESS

Table 1. Structural relationship between surfaces and regions (layers and fluids), where F denotes the foreground and B the background.

surface \ regions	$L0$	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	$L10$	$L11$	$V0$	$V1$
$S0$ – ILM	F	B	B	B	B	B	B	B	B	B	B	B	B	B
$S1$ – RNFL-GCL	F	F	B	B	B	B	B	B	B	B	B	B	B	B
$S2$ – GCL-IPL	F	F	F	B	B	B	B	B	B	B	B	B		B
$S3$ – IPL-INL	F	F	F	F	B	B	B	B	B	B	B	B		B
$S4$ – INL-OPL	F	F	F	F	F	B	B	B	B	B	B	B		B
$S5$ – OPL-HFL	F	F	F	F	F	F	B	B	B	B	B	B		B
$S6$ – BMEIS	F	F	F	F	F	F	F	B	B	B	B	B	F	B
$S7$ – IS_OSJ	F	F	F	F	F	F	F	F	B	B	B	B	F	
$S8$ – IB_OPR	F	F	F	F	F	F	F	F	F	B	B	B	F	
$S9$ – IB_RPE	F	F	F	F	F	F	F	F	F	F	B	B	F	F
$S10$ – OB_RPE	F	F	F	F	F	F	F	F	F	F	F	B	F	F

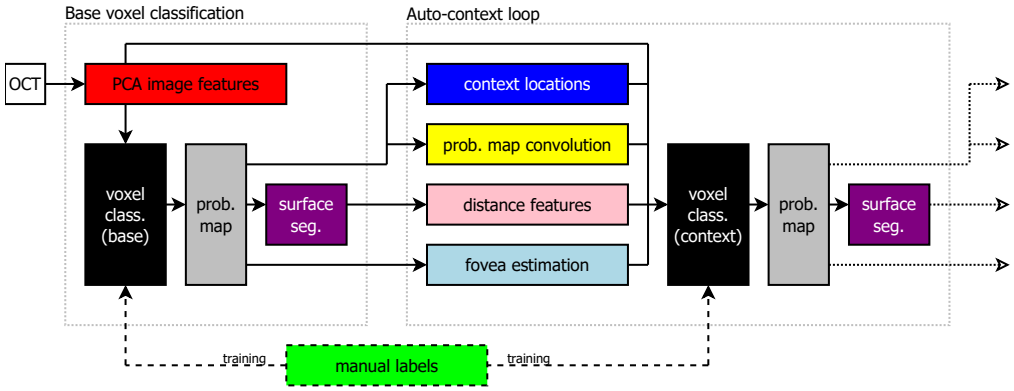

Fig. 3. Overview of the workflow of the proposed method consisting of a base voxel classification stage and a subsequent auto-context loop.

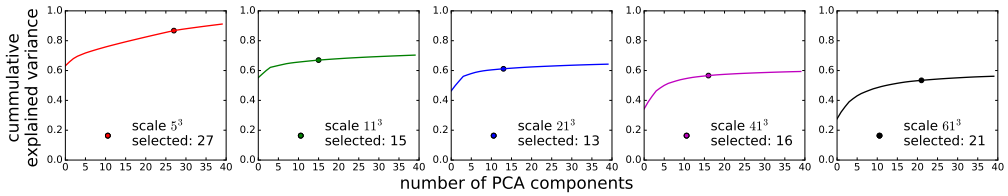

Fig. 4. The amount of variance in the original data that can be explained by a given number of PCA components. Markers show the number of components used for various scales in the proposed method.

2.3. *Base voxel classification using unsupervised representation*

In order to predict for each voxel in the OCT volume the probability that it belongs to each of the 14 regions we use a random forest machine learning algorithm [22]. The individual surfaces are then extracted from the resulting probability maps using a 3D graph-theoretic segmentation approach. A voxel-wise binary segmentation of the fluid volumes is achieved by assigning a voxel to the fluid class if the probability provided by the classifier was the maximum for that class. In the following paragraphs the method used to find suitable image feature descriptors and perform the 3D surface segmentation are described.

Biomedical Optics EXPRESS

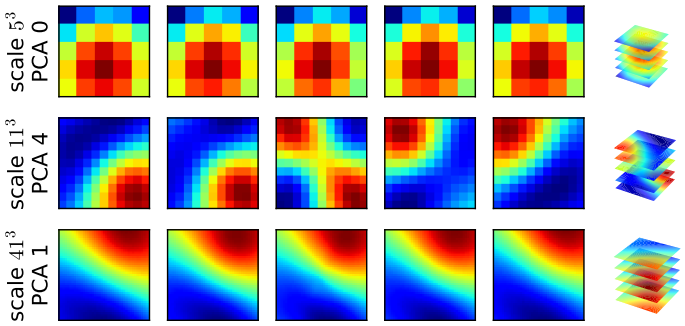

Fig. 5. PCA eigenvectors, **rows**: 3 eigenvectors at various scales, **columns**: slices of eigenvectors and the corresponding 3D representation.

PCA image features In contrast to other methods (e.g. [12]) we did not manually define image features, instead we generated convolution kernels in an unsupervised manner similar to the approach shown in [23]. We randomly pick voxels out of each OCT volume in the training set (approx. 10,000 voxels per volume, equally distributed among the regions) and extract a cubic patch around them at various scales (5^3, 11^3, 21^3, 41^3 and 61^3). For each scale an incremental principal component analysis (IPCA [24]) is then performed and the first 40 eigenvectors are computed. In order to reduce the amount of convolution kernels only the first n PCA components were used. We reconstructed the patches by using all 40 eigenvectors and then chose n so that 95% of the variability in the patches could be explained by the first n eigenvectors (Fig. 4). The selected eigenvectors are used as convolution kernels on the raw intensity data to compute the image features. As can be seen in Fig. 5 these kernels extracted from the training data resemble Gaussian and edge detection kernels.

Surface segmentation The surfaces are extracted by summing up the probabilities of the foreground regions and background regions, and using these two volumes as region costs in the graph-theoretic approach proposed in [14]. We only use this regional information and do not include any edge information or explicit feasibility constraints. Given the non-negativity of probability maps the topological ordering of the surfaces is preserved due to the foreground/background definition seen in Table 1. We do not impose constraints on the minimum/maximum distance between surfaces and we use a smoothness constraint of ±10 px between neighboring A-scans. This rather loose constraint allows the graph-search to segment along the edges of fluid regions while relying on the smoothness of the probability maps to ensure a smooth inter-layer boundary.

2.4. Auto-context loop

In order to improve the results of the method described in the previous subsection we implemented an auto-context loop. Auto-context, first proposed in [21], is an iterative approach that includes spatial context extracted from previous classifications to refine the prediction result in the next iteration.

The first iteration step is to train a classifier C_{base} on image based features f_{img} and the corresponding $labels$ (Fig. 1). The resulting probability map P_{base} is then used to extract spatial context features $f_{ctx_{base}}$. This context features are then used together with the image features to train a new classifier C_{ctx_0}. This process is then repeated for further improvements.

In order to be able to use auto-context we need a realistic probability prediction for each sample A in our training set. If we would simply use the trained classifier C_{base} to make this prediction we would get unrealistic results since A would be part of the data on which C_{base}

Biomedical Optics EXPRESS

was trained. In order to avoid this, we trained a temporary classifier C_{temp} on all samples in the training set, except A and used C_{temp} to get a prediction P_{base} for A. We repeated this step for each sample in our training set and thus got a realistic prediction for each sample. This process has to be repeated for each iteration of the auto-context loop as seen in Eq. (1).

$$\begin{aligned}
C_{base} &: train(f_{img}, labels) \rightarrow predict(C_{base}) \rightarrow extract(f_{ctx_{base}}) \\
C_{ctx_0} &: train(f_{img} + f_{ctx_{base}}, labels) \rightarrow predict(C_{ctx_0}) \rightarrow extract(f_{ctx_0}) \\
C_{ctx_1} &: train(f_{img} + f_{ctx_0}, labels) \rightarrow predict(C_{ctx_1}) \rightarrow extract(f_{ctx_1}) \\
&\vdots \\
C_{ctx_n} &: train(f_{img} + f_{ctx_{n-1}}, labels) \rightarrow predict(C_{ctx_n}) \rightarrow extract(f_{ctx_n})
\end{aligned} \tag{1}$$

The following paragraphs explain the two generic methods and the two methods specific to the retinal layer segmentation task we used to extract f_{ctx_n} from P_{n-1}.

Context locations One approach to include the probability maps of one stage into the feature vector of the next stage is to use the raw probability around each training example as a feature. In [21] this is done by extending eight rays in 45° intervals around each pixel and sampling the probability map at given intervals yielding between 5,000 and 20,000 additional context features. Instead of extending 45° rays we randomly selected 150 sample locations following a 3D Gaussian distribution (Fig. 6) around each pixel resulting in a denser sampling closer to the pixel and sparser sampling farther away.

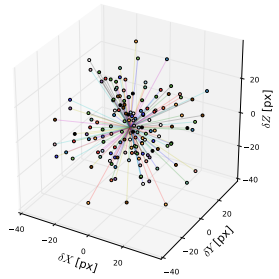

Fig. 6. 150 relative context locations sampled using a Gaussian distribution (σ = 15 px). For each voxel the samples are taken from the probability map and used as context features.

Probability map convolution Instead of only using the sampled probability, Tu et al. also used the mean of a 3 × 3 window around each sample. This operation can be expressed as a convolution of the probability map with a 3 × 3 uniform convolution kernel. Rather than using uniform kernels we compute convolution kernels using the ground truth labels by taking samples from region class A and counting how often region class B appears at each position in a small window around the sample. In Fig. 7 this counts are shown for the pair $L10$ and $L8$ in which it can be seen that the latter appears mostly a certain distance above the former. After normalization these maps are used as convolution filters on the probability maps generated by the classifier.

Distance features In addition to the general context information described above we included context information specific to the retinal layer segmentation problem. For each surface we

Biomedical Optics EXPRESS

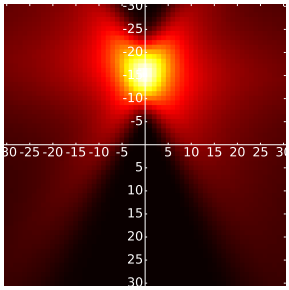

Fig. 7. Example of convolution kernel encoding the probability of R8 relative to a R10 sample at scale 61^3 px. Given that a voxel belongs to the class R8 the highest probability for finding a voxel of class R10 is approx. 15 px above it.

calculated the distance of each voxel to the corresponding foreground and background (both in 1D along the Z axis and in 3D) and the minimum 3D distance to the fluid regions.

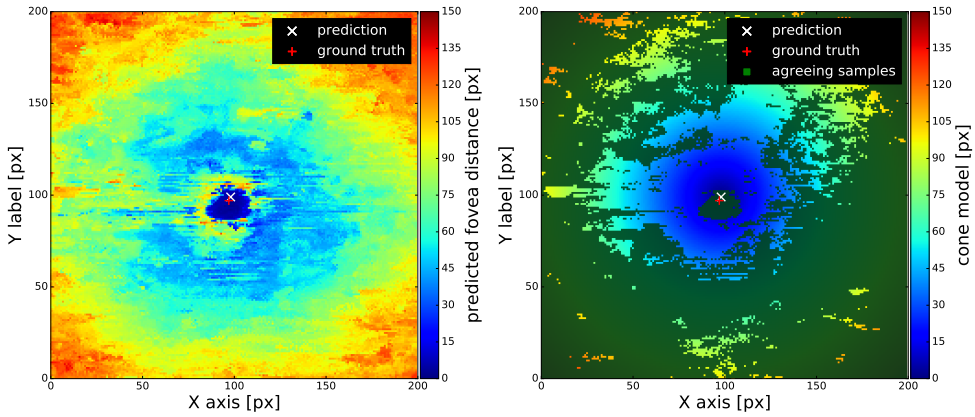

Fig. 8. Fovea position estimation. **Left:** Fovea distance prediction; **Right:** Predicted fovea position, with highlighted points in the distance map that are agreeing with the predicted position.

Fovea estimation Within the foveal region the configuration of the layers is vastly different from other regions of the retina. In order to encode this property in our training features, we compute the distance of each voxel to the fovea in the XY plane. To get an estimate of the fovea position, we calculate for each A-scan the thickness of each region and use this as a feature vector to train a random forest regressor on the XY distance to the manually annotated fovea location. On the test set we then predict the fovea distance for each A-scan resulting in a distance prediction map (Fig. 8). While the majority of the predicted distance map shows only small prediction errors, some parts are predicted incorrectly. In order to find the fovea center we used a random sample consensus (RANSAC) algorithm that is robust to such outliers. We randomly choose an A-scan as the fovea position and count the amount of points in the distance map that would agree with this fovea position (i.e. whose predicted distance matches the actual distance to the chosen fovea position). After repeating this step a number of times the fovea position for which most of the A-scans agree with is chosen as the final position.

Biomedical Optics EXPRESS

3. Evaluation and results

The proposed approach is first evaluated on a very large data set consisting of 100 OCT volumes (macula centered, $1024 \times 200 \times 200$ voxels, covering $2 \times 6 \times 6$ mm^3, Cirrus, Carl Zeiss Meditec, Inc., Dublin, CA) of patients with central and branch RVO. To generate the ground truth, the retinal scans were segmented using the method described in [14] and each B-scan was then manually corrected by trained graders resulting in a total of 20,000 annotated B-scans. In addition to the layers, the graders also annotated the IRF and SRF voxel regions as reported previously [9] and marked the location of the fovea.

In addition to the RVO data set, the algorithm was also evaluated on a publicly available data set (see [12]) containing 10 OCT volumes (macula centered, $496 \times 768 \times 61$ voxels, covering approx. $1.9 \times 8.6 \times 7.4$ mm^3, Spectralis, Heidelberg Engineering, Heidelberg, Germany) of patients with diabetic macular edema (DME). In each of the 10 volumes, manual annotation of 8 retinal layers is available in 11 B-scans within the central 6 mm. For the same set of B-scans, manual fluid annotation (containing both IRF and SRF as one class) is available as well. Thus, 110 fully annotated B-scans are used. Given the different nature of the data sets, the evaluation was performed on each set individually as described below.

Lastly, we inspect the role of the individual features used by the method on the results. We use the importance measure implemented within the random forest classifier, which relies on permuting the values of a feature and measuring how much the permutation decreases the classification accuracy of the model. Important features can then be detected as those where the permutation decreases the classification accuracy the most.

3.1. Retinal vein occlusion data set

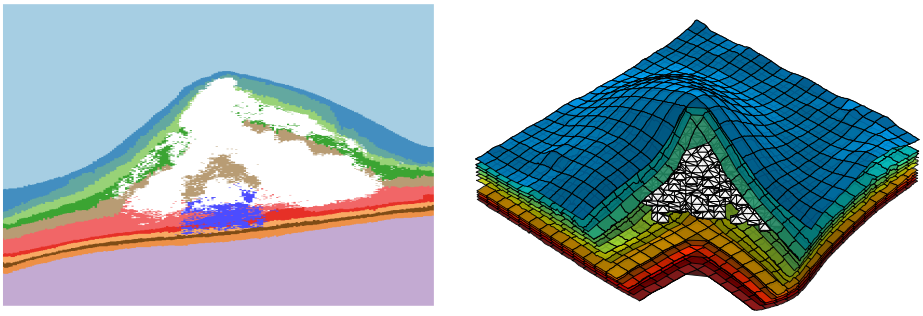

Fig. 9. Qualitative results on example image shown in Fig. 1. **Left**: Segmentation result of the proposed method after one auto-context iteration. **Right**: Corresponding 3D visualization (data was processed for visualization purposes)

For the segmentation evaluation, the data set was split into 10 subsets, one of which was used as test set, the rest alternating for training the auto-context loop as described in Section 2.4. Examples of results are shown qualitatively in Fig. 9 and Fig. 10. It can be observed that the correct layer ordering is preserved and that the resulting layers are smooth despite loose segmentation constraints.

Performance of the layer segmentation was measured using the unsigned distance from the ground truth along the Z axis (Fig. 11). It can be seen that the errors on the training data are comparable to the errors on the dedicated test set, since due to the way the auto-context classifier is trained one sample is never used to train the classifier which is used to segment it. Furthermore it can be seen that auto-context improved the segmentation result for each surface when compared to the base classifier.

Biomedical Optics EXPRESS

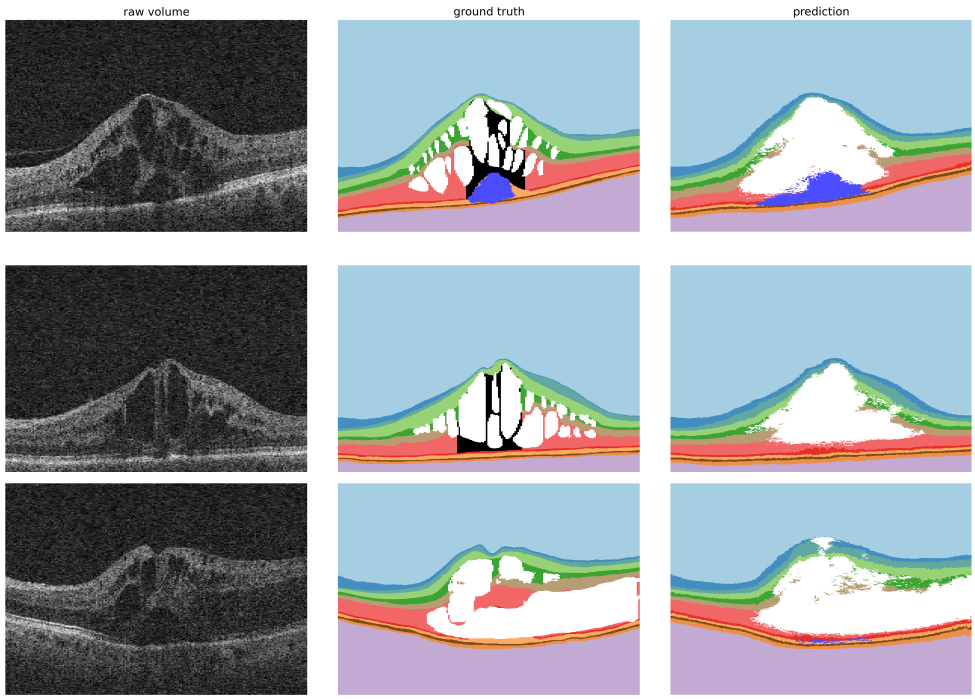

Fig. 10. Qualitative results on RVO data set. **Left:** Raw intensity image; **Middle:** Manually annotated ground truth; **Right:** Output of the proposed method.

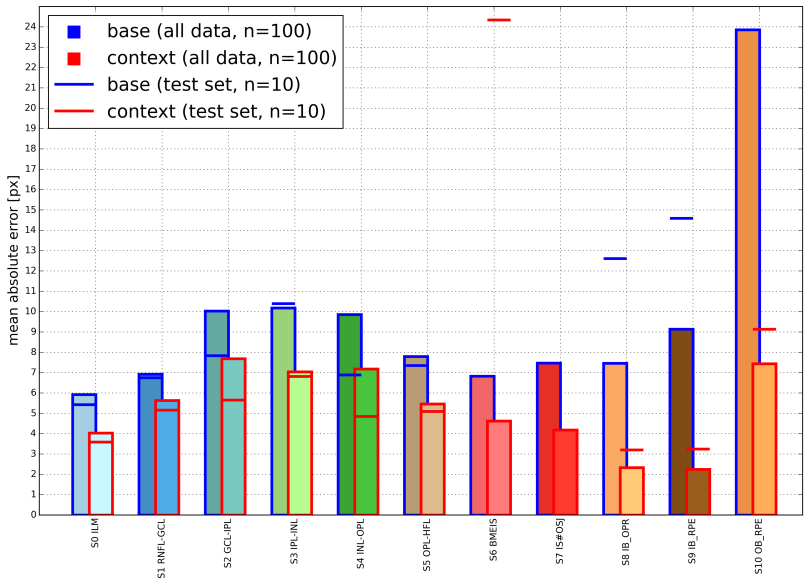

Fig. 11. Results of surface segmentation on RVO data set for training and test sets. Mean absolute error of individual surfaces without (base) and with auto-context.

Biomedical Optics EXPRESS

Table 2. Results of region segmentation on RVO data set. Dice coefficients (mean ± std) without (base) and with auto-context.

	L0	L1	L2	L3	L4	L5	L6
base	0.99±0.01	0.72±0.09	0.61±0.08	0.65±0.09	0.62±0.09	0.68±0.09	0.84±0.08
context	0.99±0.00	0.79±0.09	0.69±0.09	0.76±0.08	0.71±0.09	0.75±0.07	0.89±0.06
	L7	L8	L9	L10	L11	V0	V1
base	0.66±0.12	0.71±0.15	0.62±0.18	0.40±0.13	0.98±0.01	0.34±0.25	0.10±0.13
context	0.74±0.08	0.80±0.07	0.73±0.11	0.62±0.10	0.99±0.00	0.41±0.25	0.24±0.20

In order to evaluate the region segmentation performance, we calculated the Dice coefficient for each region as defined in Eq. (2)

$$Dice = \frac{2 * T_P}{|true| + |predicted|} = \frac{2 * T_P}{(T_P + F_N) + (T_P + F_P)} = \frac{2 * T_P}{2 * T_P + F_N + F_P} \tag{2}$$

with T_P and T_N being the true positives and negatives (i.e. the correctly classified voxels), F_P and F_N being the false positives and negatives (i.e. the incorrectly classified voxels), $|true|$ the number of voxels that belong to the region according to the ground truth and $|predicted|$ the number of voxels belonging to the region according to the algorithm. Dice is the most used metric in validating medical volume segmentations [11] and takes into account equally the false positive and the false negatives as defined by Eq. (2). The quantitative results can be seen in Table 2. Even though the coefficients for the fluid filled regions (i.e. $V0$ and $V1$) are relatively low it can be seen that the segmentation does improve due to the auto-context approach.

Finally, in order to quantitatively evaluate the fovea position estimation we computed the absolute Euclidean distance from the ground truth resulting in a mean error of $\mu = 10.0$ px, $\sigma = 5.5$ px, median= 8.4 px (results of 10-fold cross validation, Fig. 12).

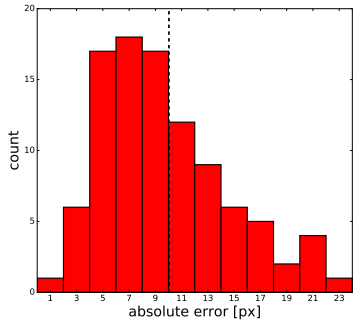

Fig. 12. Result of fovea estimation. Histogram of the absolute Euclidean distance between the predicted fovea position and the manually annotated fovea position on 100 OCT volumes in the RVO data set.

3.2. Diabetic macular edema data set

The data set provided by [12] has only 8 surfaces and one fluid class annotated, hence the surface/region definitions given in Table 1 had to be adapted accordingly (Table 3). Furthermore since the provided scans have a highly anisotropic voxel size (i.e. approx $3.9 \times 11.3 \times \mathbf{125}\ \mu m^3$ in contrast to $1.9 \times 30.2 \times \mathbf{30.2}\ \mu m^3$ in the RVO data set) the PCA image features described in section 2.3 were recomputed in 2D (see Fig. 13). Since the manual annotation was not available for every B-scan, the probability map convolution described in section 2.4 could not be performed for this evaluation. Finally, since a manual fovea position annotation is not available on this data set, the fovea position was computed in an unsupervised manner as described in [12].

Biomedical Optics EXPRESS

Table 3. Structural relationship, surfaces and regions defined for the data set provided in [12].

surface \ regions	*VITREOUS*	*NFL*	*GCL − IPL*	*INL*	*OPL*	*ONL − ISM*	*ISE*	*OS − RPE*	*BELOW − OS − RPE*	*FLUID*
ILM	F	B	B	B	B	B	B	B	B	B
NFL/GCL	F	F	B	B	B	B	B	B	B	B
IPL/INL	F	F	F	B	B	B	B	B	B	
INL/OPL	F	F	F	F	B	B	B	B	B	
OPL/ONL	F	F	F	F	F	B	B	B	B	
ISM/ISE	F	F	F	F	F	F	B	B	B	
OS/RPE	F	F	F	F	F	F	F	B	B	F
BM	F	F	F	F	F	F	F	F	B	F

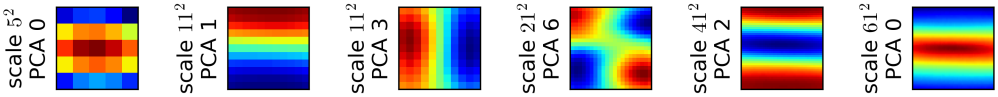

Fig. 13. 2D PCA eigenvectors computed on DME data set at various scales.

Table 4. Results of region segmentation on DME data set. Dice coefficients (mean ± std) for method without (base), with auto-context, inter-reader and the method in [12].

region	base	auto-context	inter-reader	KR + GTDP [12]
VITREOUS	0.98 ± 0.02	0.99 ± 0.00	1.00 ± 0.00	0.99 ± 0.00
NFL	0.77 ± 0.09	0.81 ± 0.05	0.86 ± 0.02	0.85 ± 0.02
GCL-IPL	0.80 ± 0.07	0.85 ± 0.04	0.89 ± 0.03	0.88 ± 0.02
INL	0.68 ± 0.09	**0.75 ± 0.05**	0.77 ± 0.04	0.73 ± 0.05
OPL	0.67 ± 0.08	**0.73 ± 0.05**	0.69 ± 0.06	0.70 ± 0.06
ONL-ISM	0.84 ± 0.05	**0.86 ± 0.04**	0.82 ± 0.06	0.80 ± 0.07
ISE	0.86 ± 0.04	**0.88 ± 0.02**	0.85 ± 0.04	0.86 ± 0.03
OS-RPE	0.68 ± 0.06	0.78 ± 0.03	0.82 ± 0.02	0.80 ± 0.04
BELOW-OS-RPE	0.99 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
FLUID	0.52 ± 0.16	**0.60 ± 0.15**	0.63 ± 0.10	0.53 ± 0.15

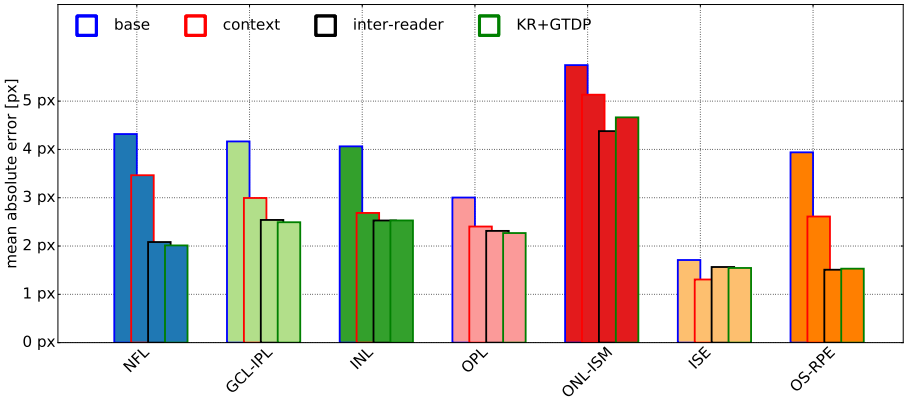

Fig. 14. Results of layer segmentation on DME data set. Mean absolute error of individual layer thicknesses for the method without (base), with auto-context, inter-reader and the method in [11].

Biomedical Optics EXPRESS

We computed the Dice coefficients for each region in the parts of the volumes where manual annotations were available. As shown in Table 4, the proposed method performs comparably to the method proposed in [12] for most regions and even outperforms it in some (in bold). In addition, we computed the mean region thickness as in [12] and compared the absolute difference to the manual annotation. It can be observed from Fig. 14 that the methods show comparable performance. Similarly to the RVO data set, an improvement due to the auto-context loop is evident in all cases.

It is interesting to note that the performance on the DME data set is consistently better than on the RVO data set even though fewer auto-context features and only 2D PCA features were used (Fig. 15). This can be attributed to the better signal to noise ratio in Spectralis scans (achieved through temporal oversampling) and the slightly less demanding segmentation task (RVO patients generally have higher volumes of SRF present than DME patients).

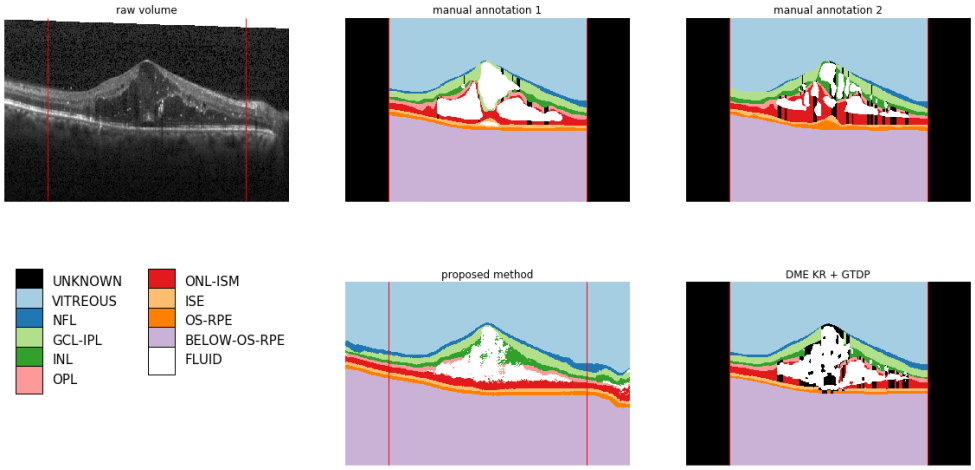

Fig. 15. Segmentation results of subject #1 in the DME data set. **Top:** Central B-scan and the manual annotations; **Bottom:** Results of the proposed method and the method presented in [12].

3.3. *Feature importance*

Results in Tables 2 and 4 show that auto-context loop consistently improves the segmentations. In the evaluation of the feature importance reported by the random forest shown in Fig. 16, it can be seen that the newly proposed context-based features have a high impact on the final result. In our analysis, the distance features show up as the most important ones, even outperforming the image based features after one auto-context loop. Similarly, the fovea distance and the probability map convolution filter also demonstrate a high impact. In contrast, the random sampling of context locations does not show consistently good results. While some highly informative locations are randomly chosen, the vast majority of sampled locations convey only limited information.

4. Discussion

Automated segmentation of retinal layers and fluid is a fundamental task for quantifying and characterizing macular edema in an objective and repeatable way. The main problem when segmenting severely diseased scans is the high variability in the shape and location of fluid filled regions. As can be seen in Fig. 10, fluid can appear in most retinal layers, but can also span

Biomedical Optics EXPRESS

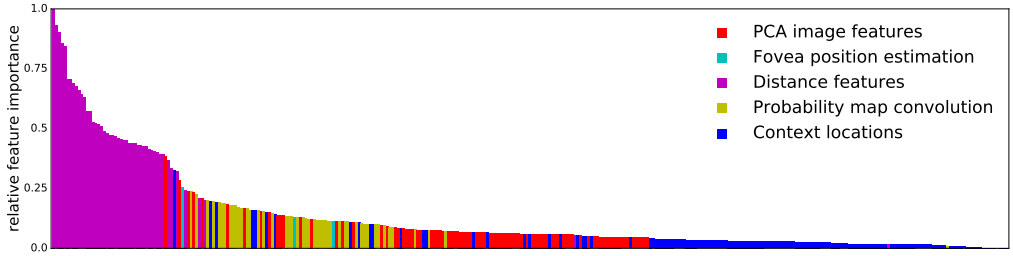

Fig. 16. Relative feature importance reported by the random forest classifier after one auto-context iteration.

multiple layers and/or be "stacked" above other fluid filled regions. This makes it very difficult to manually encode or model all the possible interactions between fluids and retinal layers.

The main contribution of this work is that the proposed method is able to learn this interaction from the training data set using a machine learning approach. In addition to simultaneously obtaining both layer and fluid segmentations the two fluid types were also differentiated. In contrast to other layer segmentation methods (e.g. [14]) the proposed method requires only high level information about the known anatomical ordering of the retinal layers (see Tables 1 and 3). As a consequence, it can be trained on data sets of different (albeit similar) pathologies without the need for manual adaptation. Similarly, the computation of the convolutional filters used as image features obtained with unsupervised representation (PCA) enables the application of the same method on different data sets with different spatial resolutions, image to noise ratios, and acquisition modes (with and without spatial oversampling), without the need of prior normalization.

Even though the proposed method achieved moderate performance in segmenting fluid regions, the segmentation still helps to "push" the surface segmentation closer to the correct position. This surface segmentation could later be used for an improved fluid segmentation - in much the same way as a pre-existing fluid segmentation could be used to improve the surface segmentation. Such a pre-existing segmentation could be used as input for the auto-context approach which enables the classifier to learn the long range spatial relations between objects in the image beyond the size of the convolution filters.

The problem of accurately segmenting retinal fluid in severely diseased cases is challenging even for qualified human readers, as noted in [12] and seen in the relatively poor inter-reader overlap shown in Table 4. Further difficulties are caused by the poor signal to noise ratio in the RVO scans used for the validation, and by the presence of large SRF volumes obscuring the retinal layers as well as the distinction between the two fluid classes. Nevertheless, while the fluid segmentation on the low-quality Cirrus scans in the RVO data set is relatively poor, the proposed method shows comparable or better performance to other methods on the Spectralis scans in the DME data set. The results could be further improved by using e.g. a more sophisticated graph-cut segmentation instead of the voxel-wise estimation, or a morphological post-processing of the segmentation results.

A limitation of our machine-learning method is the reliance on large amounts of manually labeled ground truth data in order to accurately learn the high variability present in OCT scans with severe macular edema. Nevertheless the extensive effort invested in obtaining such labeled data was rewarded with increased segmentation performance. Providing more training data especially for regions that are represented by relatively few samples in the training data (e.g. NFL in the vicinity of the fovea) would help the machine learning algorithm to learn their variance in appearance and further improve the segmentation performance. Lastly, the voxel-wise estimation of the fluid classes in the proposed method tends to oversegment the retinal fluid (Fig. 10),

Biomedical Optics EXPRESS

leading to a lower performance in the segmentation of the surrounding layers.

Regarding the computational resources, due to the way the auto-context loop is trained the algorithm requires a significant amount of processing time during the training phase (in order of a few days). For the segmentation, the algorithm requires on average 7.2 minutes for the PCA image feature extraction and 6.6 minutes per auto-context loop, on an Intel Core i7 (3770K @ 3.50GHz x 8) with 32GB RAM.

5. Conclusion

In this work we have introduced a novel fully automated 3D layer and fluid segmentation method based on unsupervised representation, auto-context and graph-theoretic segmentation. The proposed method simultaneously segments the layers and the two fluid related regions and learns their mutual interaction to aid the retinal segmentation. To the best of our knowledge, this is the first retinal layer and fluid segmentation to employ an iterative improvement approach yielding reasonable results even on highly pathologic cases. Furthermore, we have shown that the introduction of spatial context using auto-context and the domain specific context features improve the segmentation results. In particular the distances to surface segmentations of the previous iteration were shown to be important new features that improved the segmentation accuracy. Such a methodology has further potential to be adapted to other medical image segmentation problems.

The method's performance was extensively evaluated on a very large real world data set consisting of 100 fully annotated OCT volumes that show signs of severe macular edema, yielding a mean unsigned surface position error of only 5.26 ± 18.75 px and a mean overall Dice coefficient of 0.76. A second evaluation was performed on a publicly available data set, yielding a mean absolute thickness error of 2.87 ± 3.60 px and a mean overall Dice coefficient of 0.78 on the retinal regions. Such an accurate automated segmentation of retinal layers in highly pathological SD-OCT scans can be used as starting point for more accurate fluid segmentations or as an important imaging biomarker in itself.

Funding

The financial support by the Austrian Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged.

Acknowledgments

The authors would like to thank Prof. Sina Farsiu and his group at Duke University for providing the DME data set.