A Generative Model for OCT Retinal Layer Segmentation by Integrating Graph-Based Multi-surface Searching and Image Registration

Yuanjie Zheng[1], Rui Xiao[2], Yan Wang[1], and James C. Gee[1]

[1] Penn Image Computing and Science Laboratory (PICSL), Department of Radiology
[2] Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Abstract. We proposed a generative probabilistic modeling framework for automated segmentation of retinal layers from Optical Coherence Tomography (OCT) data. The objective is to learn a segmentation protocol from a collection of training images that have been manually labeled. Our model results in a novel OCT retinal layer segmentation approach which integrates algorithms of simultaneous searching of multiple interacting layer interfaces, image registration and machine learning. Different from previous work, our approach combines the benefits of constraining spatial layout of retinal layers, using a set of more robust local image descriptors, employing a mechanism for learning from manual labels and incorporating the inter-subject anatomical similarities of retina. With a set of OCT volumetric images from mutant canine retinas, we experimentally validated that our approach outperforms two state-of-the-art techniques.

1 Introduction

The major cause of adult blindness in developed countries is the progressive dysfunction and death of retinal photoreceptors [1]. Photoreceptors, located in the outer nuclear layer (ONL), function cooperatively with the retinal pigment epithelium (RPE). They catch and generate signals that are relayed to the second order neurons in the inner nuclear layer (INL) before being transmitted via ganglion cells to higher visual centers in the brain. The unique laminar organization of the vertebrate retina coupled with the development of novel non-invasive optical coherence topography (OCT) has revolutionized the ability to diagnose and follow the progression of photoreceptor diseases.

However, quantitative analysis of ONL and INL thickness throughout the entire accessible surface of the retina is rarely being done other than for preclinical or clinical investigation purposes. The reason for this limitation is due to the need to perform manual segmentation of the retinal layers, an effort that is extremely time-consuming as it can take up to one week [2] for a single imaged retina! This has clearly hampered the ability to translate such analysis from a research to a clinical setting, and prevented it to become a routine procedure.

Automated segmentation of retinal layers can help to provide an objective, quantitative and sensitive method for thickness analysis of retinal layers imaged non-invasively by OCT. Currently, commercially-available OCT imaging and analysis software are

K. Mori et al. (Eds.): MICCAI 2013, Part I, LNCS 8149, pp. 428–435, 2013.

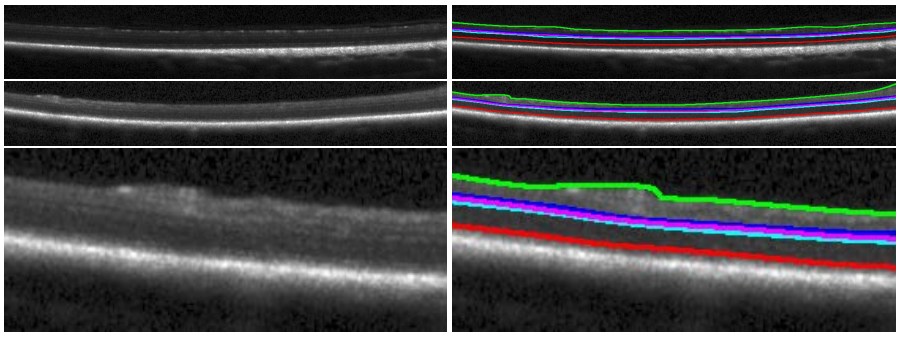

Fig. 1. Top two rows, from left to right: OCT cross-sectional scans of mutant dog retina and layer interfaces specified manually (top row) and automatically (2nd row) with the proposed algorithm, indicated with color "red" for ELM, "green" for ILM, "blue" for IPL1, "cyan" for OPL1 and "magenta" for OPL2. ELM-ILM, ELM-OPL1 and OPL2-IPL1 define the full retinal thickness, the ONL layer (thickest dark band) and the INL layer (thinner dark band), respectively. The bottom row shows an enlarged view of a local patch of the second-row images.

limited to offering measurements of either the full retinal thickness and/or that of the nerve fiber layer (NFL), incapable of segmenting other layers such as the ONL and INL. Several automated retinal layer segmentation approaches [3,4,5,6] were recently proposed to segment more retinal layers. They exploit a set of local image descriptors chosen manually or by a learning process to distinguish retinal layer's region or interfaces from other tissues, and optimally locate layers with certain optimization techniques. However, most of them are limited in accuracy and infeasible for a clinical setting due to the subtle differences between certain layers (e.g. IPL1, OPL1 and OPL2 as shown in Fig. 1) and large inter/intra-subject variations of photometric or geometric properties of OCT retinal layers.

In this paper, we propose a generative probabilistic modeling framework in order to develop automatic segmentation of OCT retinal layers. The objective is to learn a segmentation protocol from a collection of manually labeled images. Different from previous work, our framework exploits not only the discrimination ability of a set of local image descriptors learned to distinguish retinal layer interfaces from other tissues but also the anatomical similarity between the labeled image and test image (as shown in Fig. 1). Our approach combines algorithms of graph-based simultaneous searching of multiple interacting surfaces, image registration and machine learning.

This paper is novel in sense of not only a newly proposed generative model which integrates segmentation, registration and learning to more comprehensively make use of image information derived from training images, training labels and test image, but also an early attempt to apply image registration to OCT retinal layer segmentation.

2 The Generative Model

Let us use S to represent a set of layer interfaces to be automatically estimated in the test OCT image I, S_n to denote those manually delineated layers for each of the training

OCT images $\{I_n\}$, and T_n to stand for the spatial mapping field between the training image I_n and the test image I.

Our retinal layer segmentation algorithm is designed to maximize the conditional probability of $S, \{T_n\}$ given $I, \{I_n\}, \{S_n\}$, i.e. find the mode of the posterior distribution $p\left(S, T_n | I, \{I_n\}, \{S_n\}\right)$. This form of Bayesian estimation is known as a *maximum a posteriori* (MAP) estimation and written as

$$\hat{S}, \{\hat{T}_n\} = \arg \max_{S, \{T_n\}} p\left(S, \{T_n\} | I, \{I_n\}, \{S_n\}\right). \tag{1}$$

With Bayesian inference, Eq. (1) is reformulated as

$$\begin{aligned} \hat{S}, \{\hat{T}_n\} = \arg\max\nolimits_{S, \{T_n\}} & p\left(I|S\right) p\left(\{S_n\}|S, \{T_n\}\right) p(S) \\ & p\left(\{I_n\}, I | \{T_n\}, \{S_n\}, S\right) p\left(\{T_n\}\right). \end{aligned} \tag{2}$$

In Eq. (2), the first term denotes the probability of producing the test image given layer interface segmentation. The second term represents the likelihood of training labels mapped to the coordinate space of the test image, and its maximization constraints the layer interface segmentation to be adherent to the transformed manual labels. The third term means the prior probability of S. The fourth term represents the conditional likelihood related to registering the intensities of the training images and test image given their retinal layer segmentations. The last term denotes the prior probability of the transformation field of registering training image to the test image.

The MAP estimation problem of Eq. (2) is in general difficult to solve due to the need to simultaneously solve multiple unknown parameters $\{S, \{T_n\}\}$. For simplicity, we employ an iterative optimization strategy as in [7]. Apparently, the former three terms of Eq. (2) are related to retinal layer segmentation while the fourth and fifth terms are associated with retinal image registration. Following the Bayesian inference theory, the segmentation module is represented as

$$\hat{S}^k = \arg\max_{S^k} p(S^k | T_n^k(S_n), I) = \arg\max_{S} p(I|S^k) \cdot p(\{T_n^k(S_n)\}|S^k) \cdot p(S^k) \tag{3}$$

where k indexes each iterative step, and the registration module is expressed by

$$\begin{aligned} \{\hat{T}_n^{k+1}\} = \arg\max\nolimits_{\{T_n^{k+1}\}} & p(\{T_n^{k+1}\}|S^k, \{S_n\}, \{I_n\}, I) = \\ \arg\max\nolimits_{\{T_n^{k+1}\}} & p(\{I_n\}, I|\{T_n^{k+1}\}, \{S_n\}, S^k) \cdot p(\{T_n\}. \end{aligned} \tag{4}$$

The MAP problem of Eq. (2) is then resolved by iteratively solving the segmentation formulation in Eq. (3) and registration formulation in Eq. (4) until convergence or over a certain number (e.g. 4 in practice) of times.

2.1 Segmentation

We assume that the training images are independent from each other in terms of the mapping fields $\{T_n\}$ and the retinal layers' manual labels $\{S_n\}$. By removing the iteration index, Eq. (3) is rewritten as

$$\hat{S} = \arg\max_{S} p(I|S) \cdot \left(\prod_{n=1}^{N} p(\{T_n(S_n)\}|S) \right) \cdot p(S) \tag{5}$$

where N denotes the number of training images.

To define a mathematical representation of retinal layer interfaces, suppose the layer segmentation S contains L layer interfaces denoted by a set $\{\varsigma^i | 1 \leq i \leq L\}$. Similarly, the manually labeled retinal layer interfaces for S_n are represented by a set $\{\varsigma_n^i | 1 \leq i \leq L\}$. In the volumetric image $I(x, y, z)$, a layer interface ς^i is a function mapping (x, y) to z-values, i.e. $\varsigma^i(x, y) \rightarrow z$. Treating $I(x, y, z)$ as a multicolumn model which consists of a set of columns defined by their (x, y) coordinates, $\big(x, y, \varsigma^i(x, y)\big)$ is the intersection point of column (x, y) with the ith interface. As a retinal layer interface, ς^i is commonly assumed to intersect with each column of I exactly once [3].

Learn from Manual Delineations: To solve $p(I|S)$ in Eq. (5), for each of the L retinal layer interfaces, we trained an AdaBoost classifier [8] using the self-similarities (SS) image descriptors [9] computed from the training images $\{I_n\}$ and with the corresponding manual segmentations $\{S_n\}$. These trained classifiers are then applied to the test image I to get the probability $p(x, y, z, i)$ of each pixel (x, y, z) belonging to the ith retinal layer interface.

Transform Manual Delineations: In the second term of Eq. (5), $T_n(S_n)$ transforms the manual segmentation S_n of layer interface for the nth training image into the coordinate space of the test image. Suppose the shortest distance of a pixel (x, y, z) in the test image to the ith transformed layer interface is D_{ni}. Then, we set the probability of (x, y, z) belonging to the ith layer interface as $p'_{ni}(x, y, z, i) \propto \exp(-D_{ni}^2/\sigma_m^2)$. Considering all training images, we have

$$p'_i(x, y, z, i) \propto \prod_{n=1}^{N} \exp(-D_{ni}^2/\sigma_m^2), \tag{6}$$

where $p'_i(x, y, z, i)$ denotes the probability of (x, y, z) belonging to ith layer interface based on the transformed layer interfaces from all training images.

Constrain Layer Interfaces: We incorporate constraints on smoothness of layer interfaces and range of distance between layers in the third term of Eq. (5):

$$\begin{array}{r} p(S) \propto \prod_i \prod_{(x,y)} \prod_{(x',y') \in \mathcal{N}(x,y)} \exp\left(- \left(\varsigma^i(x, y) - \varsigma^i(x', y')\right)^2 / \sigma_s^2 \right) \\ \prod_{i,j \neq i} \prod_{(x,y)} \exp\left(- \left(\varsigma^i(x, y) - \varsigma^j(x, y) - m_{d(i,j)}\right)^2 / \sigma_{d(i,j)}^2 \right) \end{array} \tag{7}$$

where i and j index layer interface, $\mathcal{N}(x, y)$ denotes the neighboring pixels of (x, y), σ_s is a constant controlling the smoothness degree, $m_{d(i,j)}$ and $\sigma_{d(i,j)}^2$ are the mean and variance of the distances between the ith and jth layer interfaces.

With above equations, the probability maximization in Eq. (5) can be carried out by a minimization of the following objective function

$$\begin{array}{r} \mathcal{O}(S) = \sum_{(x,y,z)} - \ln p\left(x, y, z, S(x, y, z)\right) + c_1 \sum_{(x,y,z)} \sum_{n=1}^{N} D_{nS(x,y,z)}^2 \\ + c_2 \sum_i \sum_{(x,y)} \sum_{(x',y') \in \mathcal{N}(x,y)} \left(\varsigma^i(x, y) - \varsigma^i(x', y')\right)^2 \\ + \sum_{i,j \neq i} \sum_{(x,y)} \left(\varsigma^i(x, y) - \varsigma^j(x, y) - m_{d(i,j)}\right)^2 / \sigma_{d(i,j)}^2 \end{array} \tag{8}$$

where c_1 and c_2 are determined by σ_m in Eq. (6) and σ_s in Eq. (7), respectively.

We employ an optimization approach similar to the techniques in [3,4,6], which tries to transform the optimal and simultaneous segmentation of multiple layer interfaces into a layered-graph-theoretical problem and solve it by finding a minimum-cost s-t cut. We empirically set up the parameters of Eq. (8) as $c_1 = 3.5$, $c_2 = 2.7$, m_d/σ_d (in unit of millimicron) being $86.9/12$, $35/8$, $35/8$, and $158/90$ for layer interfaces of "ELM-OPL1", "OPL1-OPL2", "OPL2-IPL1" and "IPL1-ILM", respectively.

2.2 Registration

If we assume the registration of one training image is independent from other training images and remove the iteration index, Eq. (4) is rewritten as

$$\{\hat{T}_n\} = \arg\max_{\{T_n\}} \prod_n p(I_n, I|T_n, S_n, S) \cdot p(\{T_n\}. \tag{9}$$

To practically accomplish the maximization of Eq. (9), we ignore $p(\{T_n\}$ in order for simplicity. We first performed a global registration between each training image and test image by ignoring all segmentations and using the FSL FLIRT tool [10] with six degrees of freedom and the default parameters. Then, deformable registration was conducted using the ANTS Symmetric Normalization (SyN) algorithm [11], with the cross-correlation similarity metric (with radius 2) and a Gaussian regularizer with variance being 9. To register a training image to the test image, we ran the ANTS algorithm on each of the divided image regions by the segmented retinal layer interfaces, including the "outer-ELM", "ELM-OPL1", "OPL1-OPL2", "OPL2-IPL1" and "IPL1-ILM" regions, as shown in Fig. 1. The way we employed to get a smooth deformation field from the registration results for all divided regions is to linearly combine the velocity field outputs. After registration, reference segmentations from each of the training images were warped into the test image space.

2.3 Algorithm Overview

Our algorithm from the MAP estimation in Eq. (1) for OCT retinal layer interface segmentation is summarized as below.

Algorithm - OCT Retinal Layer Segmentation

1. Train AdaBoost classifiers with $\{I_n\}$ and $\{S_n\}$
2. Run AdaBoost classifiers on I to get $p(x, y, z, i)$ denoting each pixel's probability belonging to a layer interface
3. Minimize Eq. (8) (removing the second term) with the graph-based multi-surface searching algorithm to get an initial segmentation
4. Run FSL FLIRT tool to globally register each of $\{I_n\}$ to I
5. Divide each of $\{I_n\}$ into subregions with manual labels
6. Repeat following steps until convergence or over four times
   (a) Divide test image into subregions with current layer segmentation
   (b) Register each of $\{I_n\}$ to I with ANTS and combine with the results of FSL FLIRT
   (c) Minimize Eq. (8) with the graph-based multi-surface searching algorithm to get a refined segmentation

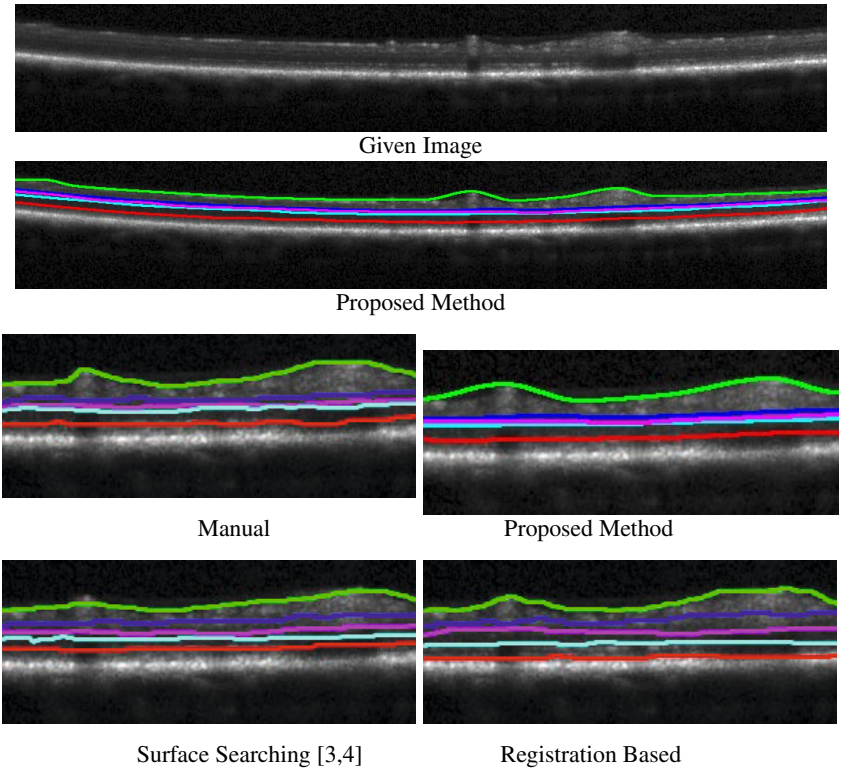

Fig. 2. Segmentation results of five retinal layer interfaces shown with different colors on one OCT cross-sectional scan of mutant dog's retina from three segmentation methods

3 Experimental Results

We employed a set of Heidelberg Engineering Spectralis OCT raw datap containing 5 volumetric images in format of ".vol", size of $1536 \times 496 \times 49$ and resolution of $5.8\mu m \times 3.9\mu m \times 123.9\mu m$. This data set was acquired from the retinas of research mutant dogs with naturally-occurring inherited retinal degeneration. A trained rater manually delineated the layer interfaces of ELM, OPL1, OPL2, IPL1 and ILM for 3 images.

We first exploited these three manually labeled images to experimentally compare the accuracies of the proposed method which integrates learning, surface searching and image registration as shown in Eq. (2), Eq. (3) and Eq. (4), a pure surface searching method similar to the techniques in [3,4] with an on-surface cost determined by image gradient as in [4] and a set of intra-/inter-surface constraints same to the ones in Eq. (7), and a registration based segmentation technique accomplished by using the FSL FLIRT tool for a global registration followed by the ANTS algorithm for a nonrigid registration as described in Sec. 2.2 and by exploiting the Majority Voting based label fusion strategy [12]. For our method and the registration based segmentation method, a leave-one-out strategy is exploited. Mean and standard deviation of the unsigned retinal layer positioning errors are computed.

Table 1. Mean±standard deviation in millimicron of the unsigned positioning errors (shortest distance of a point on an algorithm-produced interface to a manually-specified interface) of three methods on three retinal layers

Method	ILM	IPL1	OPL2	OPL1	ELM
Proposed Method	24.3±12.9	14.5±8.8	9.3±6.5	9.6 ± 5.5	18.4± 13.7
Surface Searching	27.4±14.3	25.7±19.9	24.9±18.3	24.2±19.5	18.9±14.6
Registration Based	23.8±14.5	30.1±22.3	34.4±25.0	32.9±24.2	30.3±14.8

From the results in Table. 1 and Fig. 2, we got three findings. First, the proposed method accomplishes significant improvements in detection accuracy of the layer interfaces of IPL1, OPL2 and OPL1. Second, the proposed method generates accuracies comparable with the surface searching technique and better than the image registration based approach for ELM segmentation. Third, the proposed method performs comparably with other techniques in segmentation of ILM. We found that our method achieved an average thickness accuracy improvement of 13 millimicron compared to the surface searching technique and 21 millimicron with the registration based technique, over the layers of full retina, ONL and INL. This accuracy improvement is of much more practical importance compared with the need of more computational cost by our algorithm compared with other two methods.

We also visually compared the retinal layer interface segmentation results of these three segmentation algorithms on the left two OCT images for which the manual labels are unavailable. For the proposed method, we used all the manually labeled images as the training set. Similar finding were obtained: the proposed method outperforms other techniques in segmentation of IPL1, OPL2 and OPL1 and performs comparably with other techniques on ILM and ELM segmentations.

4 Conclusion

In this paper, we proposed a generative probabilistic model for automated segmentation of retinal layers from OCT image (our model is also extendable to other segmentation tasks, e.g. [13,14]). With Bayesian theories, we proved that our model can be practically approached with existing algorithms on simultaneous searching of interacting surfaces, image registration and learning. Our model integrates the benefits of employing certain prior knowledge on the layout of retinal layer interfaces, a set of effective self-similarity descriptors, mechanism of learning from manual labels and incorporation of inter-subject similarity of retinal anatomies. We experimentally validated that our algorithm outperforms two state-of-the-art techniques in or even outside the field of OCT retinal layer segmentation.

Our future work would focus on validation with a larger data set, segmentation of other retinal layers and clinical studies on retinal degenerations.

Acknowledgement. Thanks to Dr. William A. Beltran and Dr. Gustavo D. Aguirre at UPenn Veterinary Medicine for providing the OCT data sets. Thanks to Dr. David Brainard at Department of Psychology of UPenn for his instructive suggestions. This work was made possible by support from National Institute of Health (NIH) via grant P30-EY001583.

References

1. Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M., Bhattacharya, S.S.: Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. Nature Reviews Genetics 11, 273–284 (2010)
2. Beltrana, W.A., et al.: Gene therapy rescues photoreceptor blindness in dogs and paves the way for treating human x-linked retinitis pigmentosa. PNAS 109, 2132–2137 (2012)
3. Garvin, M.K., Abramoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. IEEE Trans. Medical Imaging 28, 1436–1447 (2009)
4. Dufour, P.A., Ceklic, L., Abdillahi, H., Schroder, S., Dzanet, S.D., Wolf-Schnurrbusch, U., Kowal, J.: Graph-based multi-surface segmentation of oct data using trained hard and soft constraints. IEEE Trans. Medical Imaging (2012)
5. Yazdanpanah, A., Hamarneh, G., Smith, B.R., Sarunic, M.V.: Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. IEEE Trans. Medical Imaging 30, 484–496 (2011)
6. Xu, L., Stojkovic, B., Zhu, Y., Song, Q., Wu, X., Sonka, M., Xu, J.: Efficient algorithms for segmenting globally optimal and smooth multi-surfaces. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 208–220. Springer, Heidelberg (2011)
7. Lu, C., Chelikani, S., Papademetris, X., Knisely, J.P., Milosevic, M.F., Chen, Z., Jaffray, D.A., Staib, L.H., Duncan, J.S.: An integrated approach to segmentation and nonrigid registration for application in image-guided pelvic radiotheraphy. Medical Image Analsis 15, 772–785 (2011)
8. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
9. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: CVPR (June 2007)
10. Smith, S.M., et al.: Advances in functional and structural mr image analysis and implementation as fsl. Neuroimage 23, S208–S219 (2004)
11. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12, 26–41 (2008)
12. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46, 726–738 (2009)
13. Zhang, S., Zhan, Y., Metaxas, D.N.: Deformable segmentation via sparse representation and dictionary learning. Medical Image Analysis 16(7), 1385–1396 (2012)
14. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Towards robust and effective shape modeling: Sparse shape composition. Medical Image Analysis 16(1), 265–277 (2012)