Automated Segmentation of Intraretinal Layers from Macular Optical Coherence Tomography Images

Mona Haeker, Milan Sonka, Randy Kardon, Vinay A. Shah, Xiaodong Wu, and Michael D. Abràmoff

aDept. of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA, USA; 
bDept. of Biomedical Engineering, The University of Iowa, Iowa City, IA, USA; 
cDept. of Ophthalmology and Visual Sciences, The University of Iowa, Iowa City, IA, USA; 
dVeterans Affairs Medical Center, Iowa City, IA, USA; 
eDept. of Ophthalmology, The University of Missouri – Kansas City, MO, USA

ABSTRACT

Commercially-available optical coherence tomography (OCT) systems (e.g., Stratus OCT-3) only segment and provide thickness measurements for the total retina on scans of the macula. Since each intraretinal layer may be affected differently by disease, it is desirable to quantify the properties of each layer separately. Thus, we have developed an automated segmentation approach for the separation of the retina on (anisotropic) 3-D macular OCT scans into five layers. Each macular series consisted of six linear radial scans centered at the fovea. Repeated series (up to six, when available) were acquired for each eye and were first registered and averaged together, resulting in a composite image for each angular location. The six surfaces defining the five layers were then found on each 3-D composite image series by transforming the segmentation task into that of finding a minimum-cost closed set in a geometric graph constructed from edge/regional information and a priori-determined surface smoothness and interaction constraints. The method was applied to the macular OCT scans of 12 patients with unilateral anterior ischemic optic neuropathy (corresponding to 24 3-D composite image series). The boundaries were independently defined by two human experts on one raw scan of each eye. Using the average of the experts’ tracings as a reference standard resulted in an overall mean unsigned border positioning error of 6.7 ± 4.0 µm, with five of the six surfaces showing significantly lower mean errors than those computed between the two observers (p < 0.05, pixel size of 50 × 2 µm).

Keywords: segmentation, 3-D graph search, optical coherence tomography, ophthalmology, retina

1. INTRODUCTION

Since its first description in 1991,1 optical coherence tomography (OCT) has become an important imaging modality within the ophthalmologic community. For example, the high-resolution, cross-sectional images of the retina resulting from OCT systems are increasingly being used in the diagnosis and management of a variety of ocular diseases such as glaucoma, diabetic macular edema, and optic neuropathy. As illustrated in Fig. 1, retinal OCT images are commonly acquired in the macular region or the peripapillary region (region near the optic disk). One common scanning protocol for acquiring scans in the macular region involves the acquisition of six linear radial scans in a ‘spoke pattern’ centered at the fovea (e.g., the Fast Macular protocol on the commercially-available Stratus OCT-3, Carl Zeiss Meditec, Inc., Dublin, CA). When acquiring scans surrounding the optic disk, it is common to use a number of circular scans. An example set of six radial images from a macular OCT series can be found in Fig. 2.

In the presence of ocular disease, each intraretinal layer may be affected differently. However, even though multiple layers of the retina are identifiable on OCT images, commercially-available systems currently only segment and provide thickness measurements for one layer of the retina (i.e., the total retina on macular scans and the retinal nerve fiber layer on peripapillary scans). In order to determine the effect of ocular disease on individual intraretinal layers, the detectable layers in the retina must be segmented, and their properties (such
Figure 1. Schematic view of the macular (a–c) and circular (d–f) scanning protocols. (a) Scans in macular series on the right eye. (N = nasal, T = temporal.) (b) Scans in macular series on the left eye. (c) Visualization of acquired macular scans for one eye in 3-D. Each color represents a different 2-D scan. (d) Scans in peripapillary circular series on the right eye. (e) Scans in peripapillary circular series on the left eye. (f) Visualization of acquired circular scans for one eye in 3-D.

Figure 2. Example six raw scans in a macular scan series. Note that the colored borders correspond to those found in Fig. 1(a)–(c)
as thickness) correlated with disease state. Thus, the specific purpose of this work was to develop an automated 3-D segmentation approach for the division of the retina on macular optical coherence tomography (OCT) scans into five layers.

Fig. 3(b) shows an example of the six surfaces (labeled 1 through 6) we desired to find on each 3-D composite image. Based on histology and higher resolution OCT images (from research scanners) published in the literature,\textsuperscript{2} we assumed the surfaces roughly had the following anatomical correspondence: surface 1 corresponded to the vitreo-retinal interface (VRI), surface 2 corresponded to the separation of the retinal nerve fiber layer (NFL) above from the ganglion cell layer (GCL) below, surface 3 corresponded to the separation of the inner plexiform layer (IPL) above from the inner nuclear layer (INL) below, surface 4 corresponded to separation of the outer plexiform layer (OPL) above from the outer nuclear layer (ONL) below, surface 5 corresponded to the junction between the photoreceptor inner and outer segments (IS/OS), and surface 6 corresponded to the separation of the photoreceptor outer segment (OS) from the retinal pigment epithelium (RPE). The corresponding five layers (labeled A though E in Figure 3(b)) were thus most likely associated with the following anatomical layers: A) NFL, B) GCL + IPL, C) INL + OPL, D) ONL + IS, E) OS. It is important to note that the actual segmentation was performed in 3-D. For example, Fig. 3(c) shows a 3-D visualization of surface 3.

Our method found each surface (or set of surfaces) by transforming the 3-D segmentation problem into finding a minimum-cost closed set in a corresponding vertex-weighted geometric graph constructed from edge/regional image information and a-priori surface smoothness and interaction constraints. This type of transformation for general 3-D multiple surface segmentation problems has been previously reported by Li \textit{et al.}\textsuperscript{3} It extends a previously reported method for detecting a single optimal surface by Wu and Chen\textsuperscript{4} by adding additional edges to model interactions between surfaces. One important advantage of using this surface detection method\textsuperscript{3, 4} when compared to other previously-reported 3-D based surface segmentation methods\textsuperscript{5–7} is that it guarantees to find the three-dimensionally optimal solution with respect to the cost function.

2. METHODS

2.1. Overview

As was indicated in Fig. 1 (a–c) and Fig. 2, one macular OCT image series (using the fast macular Stratus OCT-3 protocol) consisted of six radial linear cross-sectional scans centered at the fovea. For each eye, repeated series were acquired (six, if possible), so that up to six raw scans existed at each angular location. The overall goal of the segmentation method was to determine the six surfaces defining the five retinal layers on a composite 3-D image derived from the repeated raw scans. There were two stages to the overall approach: I) the creation of a composite 3-D macular image from the raw scans and II) the determination of the six surfaces on the 3-D composite image. An overview of the data flow in the segmentation process can be found in Fig. 4.
Figure 4. Overview of segmentation steps for the data associated with one eye. First, each individual scan was aligned so that the RPE (boundary 6) was approximately horizontal in the image. Second, images from each location were registered and averaged to form a composite image. Finally, the intralayer boundaries were determined using a 3-D graph-search approach. All steps were performed automatically.
2.1.1. Overview of stage I: Creation of each 3-D composite image

The 3-D composite image associated with each eye was created in two major steps. In the first step (Fig. 5), raw scans for a particular angular location (e.g., all the vertical scans) were individually aligned so that that boundary 6 (the retinal pigment epithelium) appeared approximately straight in the aligned image. The purpose of the alignment was twofold: to aid in the final 3-D segmentation and to allow for better visualization. Each scan was aligned by first finding boundaries 1, 5, and 6 simultaneously using an optimal graph search approach similar to that used during stage II (described in more detail in later sections), but performed in 2-D. To ensure smoothness, a least-squares spline was fit to boundary 6. The columns were then translated so that this spline would be a straight line in the aligned image.

In the second step of this stage, each aligned image was registered to the first image in its location set by exhaustively searching for the best whole-pixel translation (according to the mutual information registration metric) to align each of its columns to the corresponding target image column. The position of boundary 6 determined during the first step was used as a guide to determine the range of translations to be tested for each column. The registered images in each location set were averaged together to form the composite image for that particular angular location. The purpose of averaging the images was to obtain a representative scan of that location that had a higher signal-to-noise ratio than any of the raw scans. An example of an individual scan and the corresponding 2-D composite scan is shown in Fig. 6. The set of 2-D composite images (one for each angular location) formed the 3-D composite image used in the next stage.

2.1.2. Overview of stage II: Segmentation of each 3-D composite image

In the second stage, the six surfaces were found on the 3-D composite image. As a pre-processing step, a speckle-reducing anisotropic diffusion method was applied. Surfaces 1, 5, and 6 were then simultaneously found using an optimal graph search approach (transforming the segmentation problem into finding a minimum-cost closed set in a geometric 3-D graph). After the determination of surfaces 1, 5, and 6, the remaining surfaces were found sequentially (allowing the utilization of other surface locations in the cost functions) in the following order: surface 4, surface 3, and finally, surface 2. The graph search approach guaranteed that the optimal feasible (satisfied smoothness and interaction constraints) surfaces would be found with respect to the designed cost-functions.

As the focus of this paper is on this 3-D segmentation stage, more details of the graph search approach and cost functions used will be described in the next sections. In particular, Section 2.2 will provide a more precise definition of the surface segmentation problem (the optimization problem to solve), Section 2.3 will briefly describe how the graph search was used to solve such an optimization problem, and Section 2.4 will describe the used cost functions in more detail.
Figure 6. Comparison between an individual scan and a 2-D composite scan (top and bottom of images have been cropped to aid in visualization). (a) Individual scan. (b) Composite scan.

2.2. The surface segmentation problem

The nature of the macular scans (Fig. 1(a–c)) made it natural to use a discrete cylindrical coordinate system when working with each 3-D composite image (the z-axis coincided with the intersection of the six 2-D composite scans). The coordinates of each voxel could thus be described with the triple \((r, \theta, z)\), where \(r\) reflected the distance of the voxel from the z-axis, \(\theta\) reflected the angular location of the voxel (0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, or 330), and \(z\) reflected the row of the voxel in the corresponding 2-D image. Note that with this coordinate system, voxels in the left half of each 2-D image had a different \(\theta\) value than those in the right half (for example, for the vertical 2-D scan shown in red in Fig. 1(a–c), voxels in the right half of the image had a \(\theta\) value of 90 while those in the left half had a \(\theta\) value of 270).

Each surface could be defined with a function \(f(r, \theta)\), mapping \((r, \theta)\) pairs to \(z\)-values. Furthermore, a surface was considered feasible if it satisfied certain smoothness and surface interaction constraints. In particular, a surface was considered feasible if:

- \(\theta\) smoothness constraint: \(|f(r, \theta + 30) - f(r, \theta)|\) was less than or equal to \(\Delta \theta\) for all \((r, \theta)\), \(\theta \leq 300\).
- Circularity constraint: \(|f(r, 0) - f(r, 330)|\) was less than or equal to \(\Delta \theta\) for all \(r\).
- \(r\) smoothness constraint: \(|f(r + 1, \theta) - f(r, \theta)|\) was less than or equal to \(\Delta r\) for all \((r, \theta)\).
- Constraint to connect the left and right halves of the 2-D scans together: \(|f(0, \theta) - f(0, \theta + 180)|\) was less than or equal to \(\Delta r\) for \((r, \theta)\) pairs in which \(0 \leq \theta \leq 150\).
- Surface interaction constraint for each pair of surfaces \(f_1\) and \(f_2\): The distance between the two surfaces was at least \(\delta^l\) voxels and at most \(\delta^u\) voxels (i.e., \(\delta^l \leq f_1(r, \theta) - f_2(r, \theta) \leq \delta^u\) for all \((r, \theta)\)).

In essence, the smoothness constraints required the \(z\)-values of neighboring surface points (see Fig. 7) on a particular surface to be within a specified range (given by \(\pm \Delta \theta\) or \(\pm \Delta r\)) and the surface interaction constraints required the surface \(z\)-values for a particular surface to be within a specified range of the corresponding points on the other surfaces.

Given a cost function \(c(r, \theta, z)\) that measures the unlikeness that each voxel belongs on a particular surface, the cost of a surface was defined as the summation of all voxel costs on the surface. Similarly, the cost of a set of surfaces was defined as summation of all the surface costs in the set. Consequently, the goal of the single surface detection problem (as used for finding each of surfaces 4, 3, and 2 sequentially) was to find the feasible surface with the lowest cost. The goal of the multiple surface detection problem (as used for finding surfaces 1, 5, and 6 simultaneously) was to find the set of feasible surfaces with the lowest cost.
2.3. Solving the surface segmentation problem using 3-D graph search

Each single and multiple surface problem was transformed into that of finding a minimum-cost (nonempty) closed set in a corresponding vertex-weighted geometric graph as described in. Note that a closed set was a subset of the vertices of the graph such that no directed edges left the set. One 3-D geometric graph was constructed per surface and edges in the graph were defined so that each nonempty closed set in the graph corresponded to a feasible surface. In the case of detecting multiple surfaces simultaneously, additional edges were added to model the surface interaction constraints. Furthermore, the vertex costs were assigned so that the cost of each surface corresponded to the cost of the corresponding closed set (within a constant). The minimum-cost closed set was then found by computing a minimum s-t cut in a closely-related graph.

2.4. Cost functions

Clearly the defined cost functions were an important component in determining the desired surfaces. In this work, the cost function for each surface was constructed from a linear combination of base “intuitive” cost function terms so as to satisfy expected properties of the surface. For example, it was expected that the first surface could be characterized by a combination of the following two properties: 1) the presence of an edge with a dark-to-light transition and 2) the lack of bright voxels above the surface. Correspondingly, the cost function for the first surface was defined as a normalized combination of a signed edge image (to favor the dark-to-light transition) and a cumulative image (created starting at the top of the image so as to discourage the detection of surfaces for which there were many bright pixels above the surface).

The cost functions for all of the surfaces followed the general pattern of having an edge-based term (to either favor a dark-to-light transition or a light-to-dark transition) and one or more regional-based terms (such as the cumulative image used in the cost function for surface 1). Depending on the prior knowledge of the locations of other surfaces, regional information used in this work generally was acquired from the locations illustrated in Fig. 8. Because both surrounding surfaces of each surface were often not known (surface 2 was the only surface for which the two surrounding surfaces were known) before designing its cost function, it was common to only use regional information from a limited region surrounding the surface (e.g., as in Fig. 8(a–b)).

More specifically, each of the surface cost functions was constructed from a normalized combination of a set of the following terms:

- Signed edge term (using Sobel kernel) favoring a dark-to-light transition (used for surfaces 1, 5, and 6).
- Signed edge term (using Sobel kernel) favoring a light-to-dark transition (used for surfaces 2, 3, and 4).
- Summation of pixel intensities in a limited region (Fig. 8(a)) above each potential surface voxel to encourage favoring surfaces with dark regions above surface (used for surfaces 5 and 6).

Figure 7. Schematic view of neighbor relationship for 3-D macular OCT segmentation. The edges indicate neighborhood connectivity of one “column” of z-values at a \((r, \theta)\) pair to another. For each edge shown, smoothness constraints existed between corresponding voxel z-columns for the two \((r, \theta)\) pairs connected to the edge. (a) Base graph using cylindrical coordinates. (b) Base graph using unwrapped coordinate system (as might be stored in the computer).
Figure 8. Some examples for where the image information comes from in a regional cost function term. Dark borders represent surrounding surfaces (may not be known) of the surface for which the cost function term is being defined. In cases for which an upper or lower surrounding surface does not exist (i.e., the first and last surfaces), the corresponding dark border represents the boundary of the image.

- Negated summation of pixel intensities in a limited region (Fig. 8(a)) above each potential surface voxel to encourage favoring surfaces with bright regions above surface (used for surface 2).
- Summation of pixel intensities in a limited region (Fig. 8(b)) below each potential surface voxel to encourage favoring surfaces with dark regions below (used for surface 3).
- Negated summation of pixel intensities in a limited region (Fig. 8(b)) below each potential surface voxel to encourage favoring surfaces with bright regions below (used for surfaces 5 and 6).
- Cumulative term acquired starting at the top of the image and accumulating downwards (Fig. 8(c)) to discourage finding surfaces with bright pixels above the surface (used for surface 1).
- Cumulative term acquired starting from the known boundary below and accumulating upwards (Fig. 8(d)) to discourage finding surfaces with bright pixels below the surface (used for surface 4).
- Chan-Vese\textsuperscript{10} inspired term that attempted to minimize the intensity variances surrounding the surface. \textit{A-priori} estimated means of the two regions separated by the surface were computed from a region surrounding each known surface (as shown in Fig. 8(e) with the lighter intensity region indicated by a dashed line). Because the best use of this term required the prior location of the two surrounding surfaces, only surface 2 used this term.

3. EXPERIMENTAL METHODS

The algorithm was tested on fast macular scans from 12 subjects with unilateral chronic anterior ischemic optic neuropathy. Note that the unilateral nature of the disease meant that we had data for 24 eyes, 12 of which were affected by optic neuropathy, 12 of which were not. In almost all cases (21/24 eyes), six repeated series were used to create the 3-D composite image for each eye. (Each of the remaining three eyes used fewer than six repeated series to create the 3-D composite image.) The resulting 24 3-D composite images were each comprised of 6 composite 2-D scans (144 total composite 2-D scans) of size 128 × 1024 pixels. The corresponding reported physical width and height of the 2-D raw scans (and thus also the composite scans) was 6 mm × 2 mm, resulting in a pixel size of approximately 50 \(\mu\)m (horizontally) × 2 \(\mu\)m (vertically).

One raw scan from each eye was independently traced by two human experts with the average of the two tracings being used as the reference standard. The experts did not attempt to trace borders that were not considered visible. The algorithmic result on the corresponding composite 2-D scan was converted into the coordinate system of the raw scan (undoing alignment/registration) and the mean and the maximum unsigned border positioning errors for each border were computed (the middle 30 pixels were not included to exclude the fovea). The unsigned border positioning errors were also computed using one observer as a reference standard for the other. For each border, a paired t-test was used to test for significant differences in the computed mean border positioning errors (\(p\)-values < 0.05 were considered significant). Corresponding 95\% confidence intervals (CI) were also computed.
As indicated by our quantitative results, our method performed very well overall. In almost all cases (except for the border positioning errors associated with surface 2), the algorithm performed statistically as well or better than the two observers. However, it is also important to recognize some of the limitations of using this approach. One example is that in some cases (especially for surface 2), the surface was not visible on the images. In these cases, the algorithm attempted to find the surface even if it was not present. A human observer, on the other hand, was able to indicate that the surface was not visible. Of course, having an undetectable surface does not necessarily imply that an intralayer is missing, as the ophthalmic community is still in the process of learning the precise anatomical correspondence of the visible OCT layers. Nevertheless, in such cases of an undetectable surface, it would be important for a human to be able to not use the surface determined by the algorithm.

To our knowledge, this is the first reported approach for the 3-D segmentation of intraretinal layers in OCT images\textsuperscript{11,12} and thus represents an important contribution for use in the ophthalmologic community. Because
Table 2. Summary of mean thickness differences† for 24 scans in μm.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Algo. vs. Avg. Obs.</th>
<th>Obs. 1 vs. Obs. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2*</td>
<td>5.0 ± 3.6</td>
<td>8.1 ± 3.1</td>
</tr>
<tr>
<td>1–3*</td>
<td>3.7 ± 3.2</td>
<td>10.6 ± 4.1</td>
</tr>
<tr>
<td>1–4</td>
<td>6.5 ± 7.2</td>
<td>18.8 ± 9.1</td>
</tr>
<tr>
<td>1–5</td>
<td>2.1 ± 1.8</td>
<td>3.1 ± 2.9</td>
</tr>
<tr>
<td>1–6</td>
<td>4.9 ± 3.2</td>
<td>6.0 ± 4.6</td>
</tr>
</tbody>
</table>

† Mean ± SD. For each boundary, thicknesses were not computed for the middle 30 pixels (out of 128) to exclude the fovea.
* Differences were not computed for those scans in which the lower boundary defining the layer was determined to not be visible by at least one expert.

Figure 9. Bar chart of mean thickness differences (error bars reflect standard deviations). Note that the difference between the algorithm and the reference standard was smaller than that between the two observers.

Figure 10. An example result representing typical performance in which all border position errors were close to the reported mean values. (a) Composite image. (b) Composite image with labeled borders.
the underlying graph search is a 3-D approach, it should also be applicable to more densely acquired 3-D OCT images, such as those resulting from Fourier-domain OCT systems (not yet commercially available).

More generally, this work also reflects the first time that the 3-D graph search approach has had to be applied to an application needing to find so many layers. Theoretically, the graph search would be capable of simultaneously finding as many layers as desired. However, when finding multiple layers simultaneously, the cost functions for all surfaces must be specified upfront. We have found that some cost functions can be better designed by allowing the use of previously found surface positions. Thus we approached this by simultaneously finding three “easier” surfaces first, and then sequentially finding the remaining surfaces in an effort to best utilize a priori-information in the cost function design.

In summary, we have presented an automated approach for the segmentation of intraretinal layers on macular OCT images, thus enabling the separate computation of individual layer properties, such as thickness. Having separate layer properties will be especially important in cases in which the individual layers are affected differently (e.g., one layer may thin due to neuron loss, while another one may thicken due to the presence of fluid). In addition, it will aid in clinical studies designed to pinpoint which layers are actually affected during disease processes.

6. ACKNOWLEDGEMENTS

This work was supported, in part, by NIH-NIBIB grant R01-EB004640-01, NIH-NEI grant R01-EY017066, a grant from the Veterans Administration (Merit Review and Rehabilitation Division), and an unrestricted grant from Research to Prevent Blindness, New York, NY.

REFERENCES