Retinal Thickness Measurements From Optical Coherence Tomography Using a Markov Boundary Model

Dara Koozekanani, *Student Member, IEEE*, Kim Boyer*, *Senior Member, IEEE*, and Cynthia Roberts

***Abstract*—We present a system for detecting retinal boundaries in optical coherence tomography (OCT) B-scans. OCT is a relatively new imaging modality giving cross-sectional images that are qualitatively similar to ultrasound. However, the axial resolution with OCT is much higher, on the order of 10 μm. Objective, quantitative measures of retinal thickness may be made from OCT images. Knowledge of retinal thickness is important in the evaluation and treatment of many ocular diseases. The boundary-detection system presented here uses a one-dimensional edge-detection kernel to yield edge primitives. These edge primitives are rated, selected, and organized to form a coherent boundary structure by use of a Markov model of retinal boundaries as detected by OCT. Qualitatively, the boundaries detected by the automated system generally agreed extremely well with the true retinal structure for the vast majority of OCT images. Only one of the 1450 evaluation images caused the algorithm to fail. A quantitative evaluation of the retinal boundaries was performed as well, using the clinical application of automatic retinal thickness determination. Retinal thickness measurements derived from the algorithm's results were compared with thickness measurements from manually corrected boundaries for 1450 test images. The algorithm's thickness measurements over a 1-mm region near the fovea differed from the corrected thickness measurements by less than 10 μm for 74% of the images and by less than 25 μm (10% of normal retinal thickness) for 98.4% of the images. These errors are near the machine's resolution limit and still well below clinical significance. Current, standard clinical practice involves a qualitative, visual assessment of retinal thickness. A robust, quantitatively accurate system such as ours can be expected to improve patient care.**

***Index Terms*—Boundary detection, edge detection, Markov models, ophthalmology, optical coherence tomography, perceptual organization, retina.**

I. Introduction

We present a new system to automatically detect the retinal boundaries in optical coherence tomography (OCT) images of the eye. OCT is a new imaging modality, analogous to ultrasound, that obtains cross-sectional images through tissues and materials. OCT offers higher resolution than high-frequency ultrasound and can be used to image the retina. These images can either be used to qualitatively assess retinal features and pathologies or to objectively make quantitative measurements. To this end, we developed our system to facilitate the measurement of retinal thickness from OCT images. Retinal thickness measurements are clinically important because they may be used to diagnose diseases and determine the appropriate treatment. For OCT to become a clinically practical tool for determining retinal thickness, a robust and accurate measurement system must be developed. With such a system, the accuracy and precision of OCT will allow improved monitoring of patients, earlier detection of pathology, and more-precise treatment protocols. To measure thickness using OCT, we must first detect the retinal boundaries within the images. The boundary-detection system we present here is demonstrably more robust than that built into the Humphrey 2000 OCT system, which, to the best of our knowledge, is the only commercially available system in the US.

We have organized the paper as follows. This section, continues with a discussion of the relevant ophthalmology and prior work; Section II explains the general theory behind OCT and gives details about the Humphrey 2000; Section III outlines the assumptions and mathematical development, Section IV explains the algorithm itself, Section V develops our method of evaluation, Section VI presents the results of extensive testing, and Section VII offers some concluding remarks.

A. Ophthalmology

So that the reader will understand the clinical significance of the problem at hand, we will briefly overview the relevant ophthalmology. The posterior portion of the eye is a hollow sphere filled with the clear, jelly-like vitreous humor and the retina is a thin film of nervous tissue lining the inner surface of the posterior ocular wall [1]. The eye's radial geometry makes the terms *inner* (toward the center of the eye) and *outer* useful for orientation within the retina. Just outside the retina lies the choroid, a tissue whose dense network of blood vessels renders it opaque to infrared light. The fovea is a small depression in the retinal surface, at the location corresponding to the center of the visual field.

Many diseases affecting vision cause changes in retinal thickness, with macular edema being a prominent example. Macular edema is the primary cause of vision loss in diabetics [2] and tracking retinal thickness changes over time is important in determining the appropriate treatment [3]. Currently, the slit lamp

Manuscript received October 30, 2000; revised June 23, 2001. This work was supported by the Whitaker Foundation and Sunbury Lions. The Associate Editor responsible for coordinating the rewiew of this paper and recommending its publication was M. W. Vannier. *Asterisk indicates corresponding author.*

D. Koozekanani is with the Biomedical Engineering Program and Signal Analysis & Machine Perception Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210-1272 USA.

*K. Boyer is with the Biomedical Engineering Program and Signal Analysis & Machine Perception Laboratory, Department of Electrical Engineering, 205 Dreese Laboratory, 2015 Neil Avenue, The Ohio State University, Columbus, OH 43210-1272 USA (e-mail: kim@ee.eng.ohio.edu).

C. Roberts is with the Biomedical Engineering Program and Department of Ophthalmology, The Ohio State University, Columbus, OH 43210-1272 USA.

Publisher Item Identifier S 0278-0062(01)08662-1.

biomicroscope is the most commonly used clinical device for assessing retinal thickness. With it, the ophthalmologist views a magnified, stereo image of the retina illuminated obliquely by a narrow slit of light and subjectively assesses thickening. Though thickening above a certain threshold is readily detected based on gross appearance, there is no measuring scale for precise quantification. This lack of quantification makes it difficult to track changes over time. OCT's quantitative nature permits accurate and precise measurements of patient changes.

B. Prior Work

A very preliminary version of this work was presented at the *2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* [4]. To the best of our knowledge, only one other computer vision paper has been published on OCT images to date. This work, by George *et al.* [5], used a dual threshold to segment the retina and the choroid from OCT images. Unfortunately, very little information is available about this system. However, that which has been presented suggests it is a fairly simple system whose output is rougher than that of the system presented here.

Though there is a dearth of articles related specifically to OCT, one can benefit from looking at articles for other imaging modalities, particularly ultrasound. Ultrasound images are qualitatively similar to OCT images in appearance, with a similar degree of noise. Thus, they present some of the same challenges for feature extraction. Many ultrasound articles report the use of snakes [6]–[8]. Lefebvre *et al.* [7] use a self initializing snake to detect the boundaries of the heel bone. The initialization is performed by matching the image characteristics with known qualities of the bone. Wang and Li [8] used snakes with variable internal energy to detect muscle boundaries within loin images, where the snakes are fit not to the strength of the edges, but to their *saliency*. Here, the term *saliency* is a concept developed by Sha'ashua and Ullman [9].

A variation of snakes that has been used to find feature boundaries in medical images is a level set approach [10], [11]. Here, the active contour within the image plane is viewed as a level set of some function, f, of two dimensions (generally, $f = 0$). The *general* problem of detecting contours within noisy images has been addressed not only by snakes [12], but by improved low-level edge detectors [13], [14] as well. In another technique, Aarnink [15] adjusted the edge-detector scale based on local image variability. Statistical approaches have been proposed, such as Bayesian decision making based on local variations in mean intensity [16] and detection of edges using finite mixture distribution analysis [17]. Thune *et al.* [17] analyzed the individual A-scans of an ultrasound image to detect the boundaries of the kidney; our boundary detector also analyzes individual A-scans.

The mathematical model used in this article is an autoregressive model, which is a type of Markov model. Many articles report the use of autoregressive models [18]–[21]. In most of these cases, the autoregressive models are used to characterize the boundaries of segmented objects within images, for the purpose of object identification. However, Zhou and Pycock's work [21] is different in that the autoregressive model is used to create and evaluate possible boundary contours of epithelial cells in microscope slide preparations. Different possible boundary initializations are chosen for each cell and based on those initializations, the boundary is effectively "grown" along equally spaced, radial lines extending from the center of each cell. At each line, the model is used to pick the most likely cell boundary from among the edges within the image. Finally, the model is again used to choose the most probable final boundary determination for each cell.

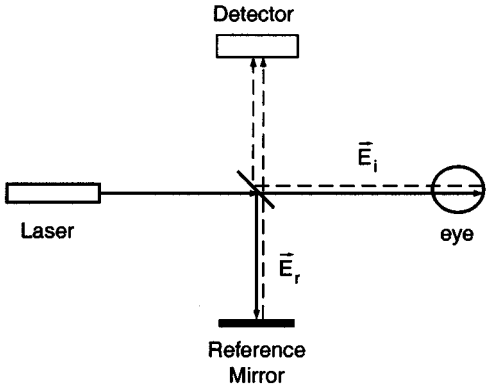

Fig. 1. Schematic drawing of the OCT emphasizesg how it is essentially a Michelson interferometer. The outgoing light paths are solid lines, while reflected light is drawn as dashed lines.

II. Optical Coherence Tomography

A. General OCT Theory

An OCT system is essentially a Michelson interferometer; the two light paths are called the reference path and the imaging path and the subject's eye terminates the imaging path [22]. Fig. 1 illustrates this concept with a schematic drawing of an OCT system. In the figure, the reflected light is represented by the electric field vectors, $\vec{E}_i$ in the imaging path and $\vec{E}_r$ in the reference path. Solid lines depict outgoing light and dashed lines depict reflected light. To understand OCT operation, one should first imagine the eyeball in Fig. 1 replaced by a mirror located exactly 1 m from the beam splitter. The reference mirror in this thought experiment is initially placed .5 m from the beam splitter and then slowly moved outwards to a distance of 1.5 m.

The average signal power at the detector, S, can be modeled as the real portion of the cross-correlation function between the reflected light from the imaging path and the reference path [23]. The derivation is as follows. For a fixed reference mirror position d, the instantaneous power at the detector $S(t)$ is, for the general case

$$S(t) = \left\| \vec{E}_i + \vec{E}_r \right\|^2 \tag{1}$$

We can simplify the notation by assuming that $\vec{E}_i$ and $\vec{E}_r$ both lie within the same plane (i.e., the light is linearly polarized in the same plane); thus, we can treat $\vec{E}_i$ and $\vec{E}_r$ as scalars, yielding

$$S(t) = E_r^2 + E_i^2 + 2E_iE_r \tag{2}$$

where E_i and E_r are the magnitudes of $\vec{E}_i$ and $\vec{E}_r$, respectively. As a final simplification, if we consider the light to be at a single frequency, we can say that

$$S(t) = (E_i + E_r)(E_i + E_r)^* \tag{3}$$

where E_i and E_r are the complex phasor representations of the two light beams and the asterisk (*) represents complex conjugation. Equation (3) can be expanded to

$$\begin{aligned} S(t) &= |E_i|^2 + |E_r|^2 + E_i E_r^* + E_r E_i^* \\ &= |E_i|^2 + |E_r|^2 + 2\mathrm{Re}\,(E_i E_r^*) \end{aligned} \tag{4}$$

and the time average taken to yield

$$\langle S(t)\rangle = \langle |E_i|^2\rangle + \langle |E_r|^2\rangle + 2\mathrm{Re}\,(\langle E_i E_r^*\rangle)\,. \tag{5}$$

If we now allow d to vary, we note that E_r and, thus, S vary in response. However, E_i does not vary with d and always equals $E_i(0)$ and so $E_i E_r^*$ can be written as $E_i(0)\,E_r^*(d)$. Also, the average power of E_i and E_r are assumed to be constant and so

$$S(d) = c + 2\mathrm{Re}\,(\langle E_i\,(0)\,E_r^*\,(d)\rangle) \tag{6}$$

where c represents a constant offset. The rightmost term in (6) is the definition of a cross correlation [24]. In our thought experiment, a mirror terminates both light paths and so the cross correlation is the autocorrelation function of the light source.

An ideal laser light source has energy at only one frequency and a purely sinusoidal autocorrelation function. Almost all OCT systems reported in the literature use a superluminescent diode (SLD) for their light source; the exceptions are some recently developed, ultrahigh resolution systems using femtosecond-pulsed lasers. An SLD power spectral density is broader than that of a laser and resembles a Gaussian with a peak at the center frequency. In our example, then, the resulting autocorrelation function resembles a sinusoid windowed by a Gaussian curve at the reference mirror position of 1 m. SLD sources presented in the literature typically have half-power widths on the order of 14 μm [25], as illustrated in Fig. 2. The OCT filters the detector signal so that only the low-frequency envelope is seen; thus, its output is a Gaussian curve centered at the mirror position. From the peak location, one can infer the location of the interface which created it; the width of the autocorrelation function determines the resolution.

To image a spot on the retina, the SLD beam stays fixed at one retinal location while the reference mirror scans through some distance. The resulting $S(d)$ is called an *A-scan*. Unlike a single mirror, the retina has many reflecting interfaces produced by impedance mismatches between adjacent layers of tissue and each one creates a peak in $S(d)$ centered at its location. The A-scan value for any given d is the sum of signals from each interface. The OCT then obtains A-scans at a series of adjacent locations, usually arranged to form a circle or a line. The A-scans are then aligned to form a two-dimensional (2-D), cross-sectional view called a *B-scan*, where each B-scan column corresponds to an individual A-scan.

OCT images then, like ultrasound images, are made up of A-scans and represent reflectivity. The noise in both modalities is also qualitatively similar. In Figs. 3 and 4, we present sample

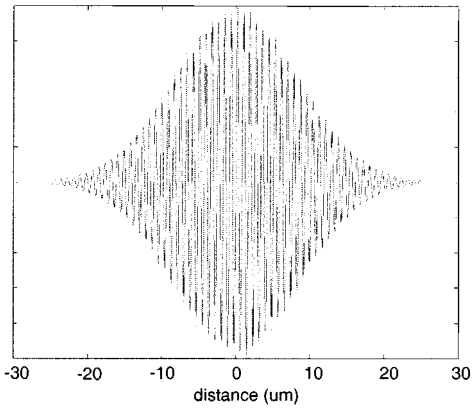

Fig. 2. Sinusoid modulated by a Gaussian curve to show conceptually what the autocorrelation function looks like. The width of the Gaussian and the period of the sinusoid match the theoretical parameters of our OCT instrument.

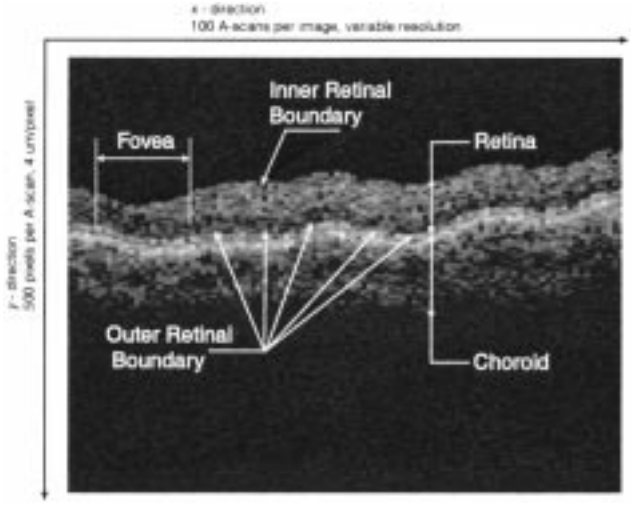

Fig. 3. Sample OCT scan from the set of training images with anatomical features labeled.

OCT scans and in Fig. 3 we have labeled the inner and outer retinal boundaries along with other retinal landmarks. Also in Fig. 3, the directions along which A-scans are obtained and then stacked are labeled. Fig. 5 plots a sample A-scan. Figs. 4 and 5 are from the same subject, while Fig. 3 is from another subject.

Though the general retinal shape is consistent between scans, the details and exact shapes may differ substantially, even within the same subject. These differences are due mainly to noise and eye motion during scan acquisition and do not represent actual retinal topography. The normal retinal surface, in fact, is smooth, such as in Fig. 4(B). The crenelations and bumps in Figs. 3 and 4(A) are artifacts caused by subject motion. These artifacts, if severe, could be a challenge for any algorithm trying to identify continuous, coherent retinal boundaries. Furthermore, note that Fig. 4(B) is much dimmer than the others owing to a *shadowing* effect caused by subject's pupil margins blocking the SLD beam, particularly on the left-hand side of the B-scan. This can cause the scan intensity and, hence, the retinal boundaries, to be obscured by the background noise. The speckle noise is prominent and the contrast between the bright and dark specks varies with the local image intensity, suggesting a multiplicative nature. This noise blurs the retinal

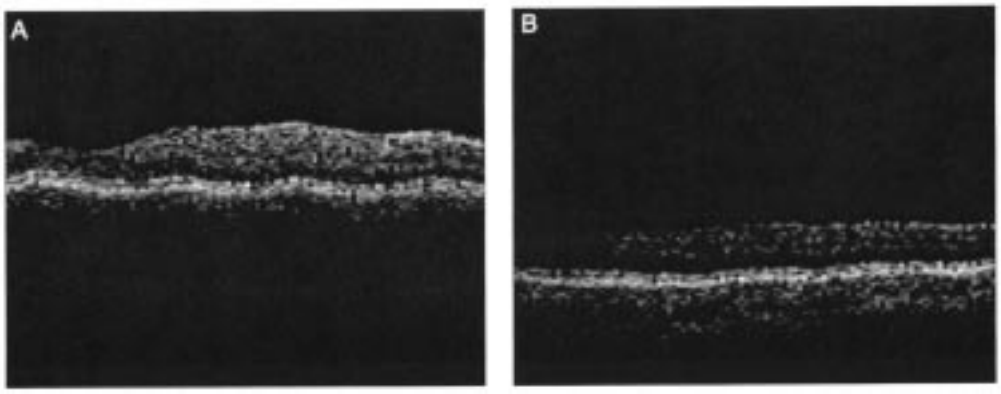

Fig. 4. Another OCT scan demonstrates the possible variations. (A) and (B) Different scans from the same person (Fig. 3 is from a different person). Note that in (A) and (B), the scans were set to the same trajectories; the differences in appearance are due to noise, differences in subject motion, and differences in pupil alignment with the OCT.

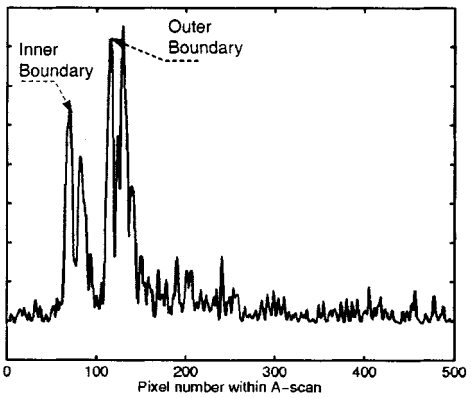

Fig. 5. Sample A-scan from Fig. 4(A) without any filtering. The constant offset is evident, as is the noise. The inner and outer retinal boundaries have been labeled.

boundaries, particularly the outer retinal boundary near the fovea. Thus, the noise challenges edge detectors, both because it obscures the retinal boundaries and because edge detectors typically assume additive Gaussian noise. Finally, there is no clear agreement in the literature as to which particular image structure represents the outer retinal boundary. Fortunately, for monitoring thickness changes in a patient from a series of OCT scans, an automatic boundary detector need not mark a specific edge in OCT images. Instead, the detector must consistently mark the *same* edge in every image so that thickness *changes* may be detected; repeatability is the primary objective.

B. The Humphrey 2000 OCT System

The Humphrey 2000 OCT system is currently the only commercially available OCT system in the United States. Much of the early research in the literature was performed on custom, prototype systems, but now work based on the Humphrey system is becoming more common. All images in this article are from a Humphrey system. The Humphrey SLD has center wavelength of 850 nm [25] and an autocorrelation function half-power width of 14 μm in air and 10 μm in the eye. The Humphrey OCT always obtains 500 pixels/A-scan and 100 A-scans/B-scan. For each A-scan, the reference mirror is moved through a distance of 2 mm. Thus, the A-scan resolution is *always* 4 μm/pixel, while the separation between A-scans varies with the total length of the scan line. Note that the pixels in the A-scan direction are oversampled, as the SLD half-power width is 10 μm in the eye.

The Humphrey OCT system displays B-scans using a pseudocolor map with 172 colors, where the colors range through (by descending signal intensity) white, red, orange, yellow, green, blue, violet, and black. With the exception of those figures that we explicitly identify as coming from the Humphrey, the images we display are *grayscale* representations of the actual interference signal intensities; they are *not* grayscale renditions of the pseudocolor maps. The Humphrey system stores the raw data for every B-scan in a disk file that we export and read as a 500 $\times$ 100 element matrix, with each A-scan corresponding to one column. The actual interference signal intensities range from zero to 1600; using the raw data avoids the quantization down to the 172 levels in the pseudocolor images.

The system comes from the manufacturer with software tools for measuring retinal thickness. However, there is no literature describing precisely how the system works. The measured retinal thickness is displayed as a plot of thickness versus retinal position, along with an image of the OCT scan and the detected retinal boundaries. Neither the image of the thickness plot nor the detected retinal boundary positions can be directly exported; instead, a screen capture must be used. One such screen capture is presented in Fig. 6. The upper left quadrant displays the OCT image with the retinal boundaries as determined by the machine marked in white. The bottom left shows the plot of retinal thickness. The upper right shows a captured frame from a video camera that images the subject's retina as it is scanned by the SLD and the lower right gives patient information and scan parameters.

The Humphrey system often gives reasonable boundary determinations for the images; examples of these are in Figs. 6 and 7, where we have simply shown the images of the OCT scans with the output boundaries displayed. However, the Humphrey boundary determinations may also be wildly erroneous; Fig. 8 presents a severe case. In those images, the Humphrey system has placed the boundaries well off the retina and within the vitreous humor. Occasionally, *no* detected boundary is displayed. The errors are not always so dramatic, of course, but they frequently occur. Examples of these more typical errors are in

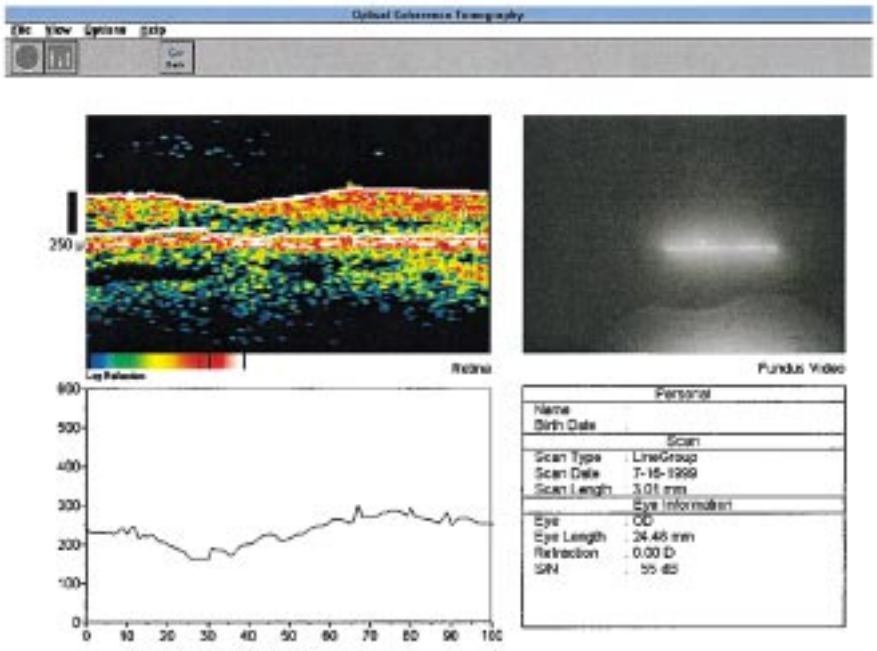

Fig. 6. Sample screen capture for the Humphrey system shows the information that is given to the user.

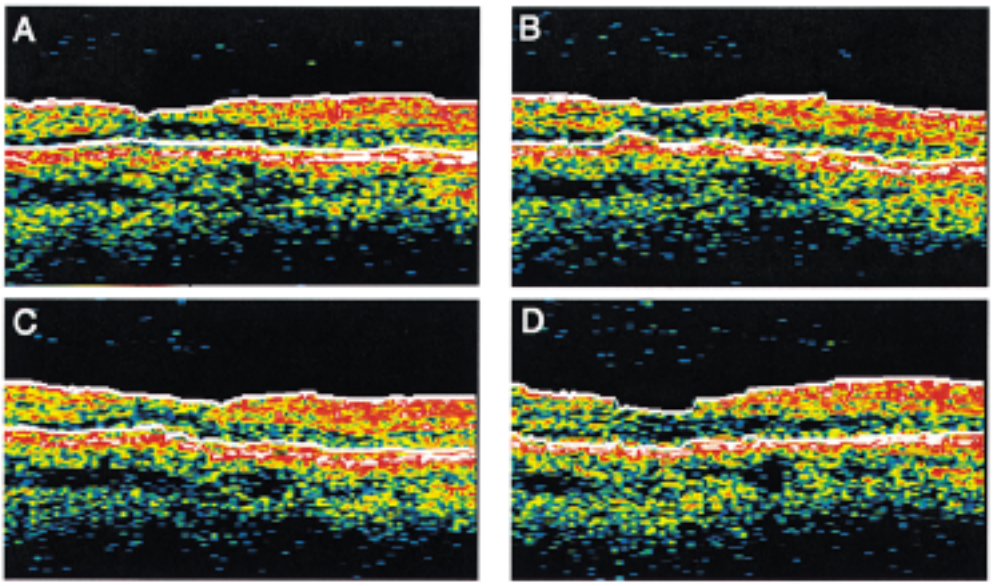

Fig. 7. More examples of acceptable performance of the Humphrey boundary-detection system.

Fig. 9. These errors are large enough to render the measurements useless. Clearly, a more reliable and automatic measurement technique is needed.

III. Theory and Models

In designing a retinal boundary detector, we used retinal anatomy and the principles of OCT operation to make various assumptions about the image boundary characteristics. To begin, the normal retina has smooth boundaries without discontinuities or gaps and, within OCT images, the inner boundary is always above the outer boundary. However, it takes about 1-s to acquire all 100 A-scans, long enough for involuntary eye motion to occur during the process. Patient motion can cause artifacts ranging from undulations to apparent breaks in the retinal image, examples of which are in Figs. 10–12. These motion artifacts can create variations in the boundary positions that do not reflect the actual retinal topography. In fact, with the exception of the foveal region, the actual curvature of the retinal surface is much lower than the artifacts caused by motion. The OCT technician may save such scans and so the boundary detector must be able to cope with them despite the assumption of retinal smoothness. However, images featuring very large dislocations are difficult even for humans to interpret and so they are routinely rejected and need not be considered.

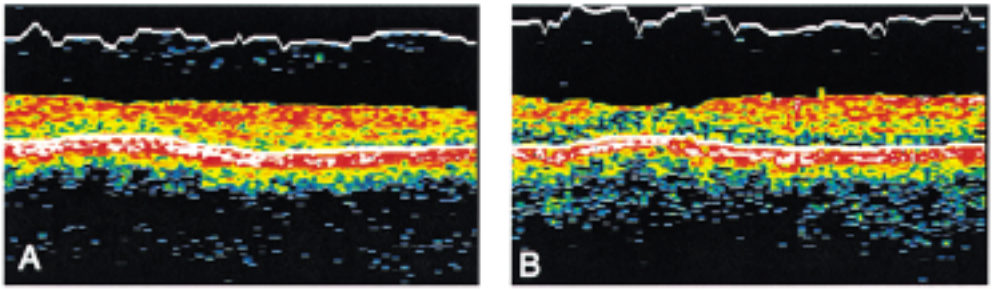

Fig. 8. Examples where the Humphrey boundary-detection system completely failed.

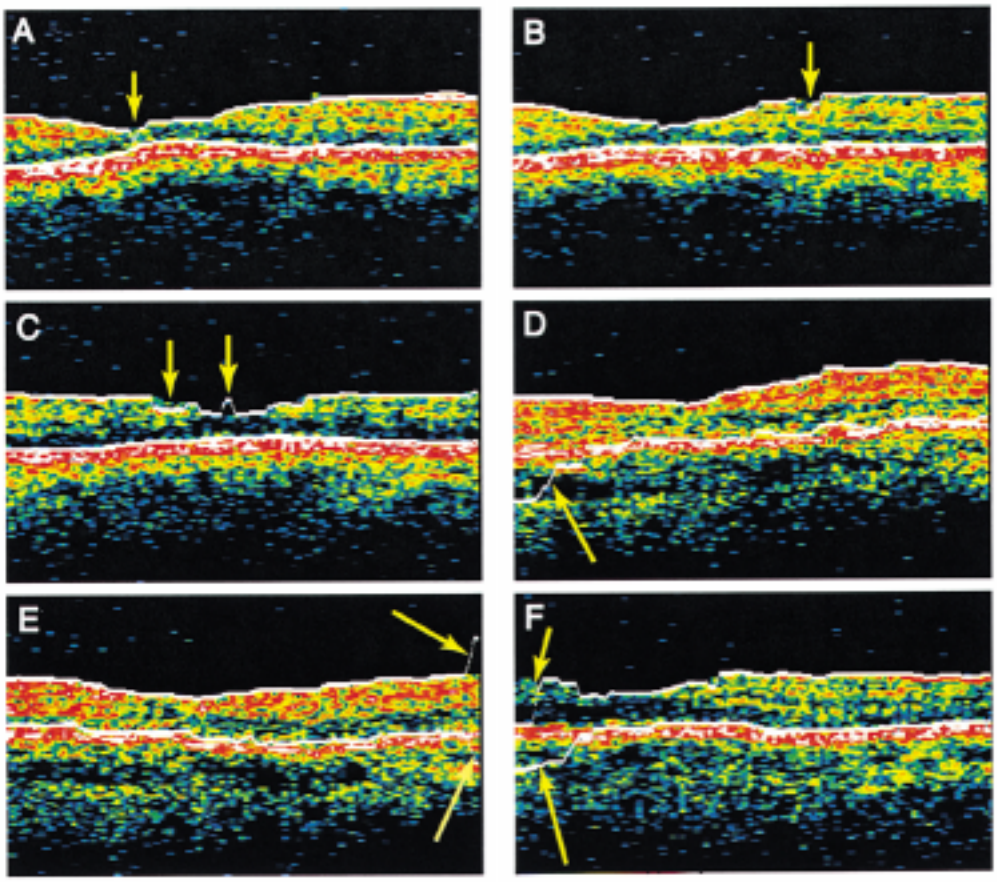

Fig. 9. Examples of errors (marked by arrows) that typically occur with the Humphrey boundary-detection system. These errors are large enough to render the measurements useless.

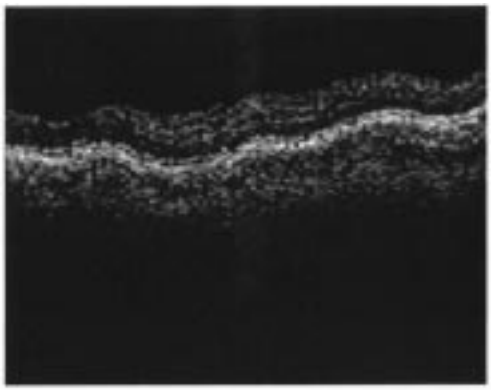

Fig. 10. Sample OCT scan shows the undulations which may be present. These undulations may be the result of tiny, involuntary eye motions.

Because the OCT acquires each A-scan individually, we apply one-dimensional (1-D) edge detection to each image column individually, similar to Thune *et al.* [17]. In fact, the multiplicative speckle noise and breaks in the images make the use of 2-D edge detectors problematic. Indeed, the spatial coherence assumptions used in the design of most 2-D, low-level vision operators, such as edge detectors, do not strictly apply in the row direction for OCT images. In particular, abrupt changes often occur between neighboring columns, caused by motion or noise. These changes can confound the output of 2-D kernels such as the Marr–Hildreth [26] by creating very strong but spurious vertical edges that obscure or distort the detected edges corresponding to the actual retinal boundaries. Finally, each A-scan penetrates both retinal boundaries and so we assume that every image column intersects exactly two boundaries.

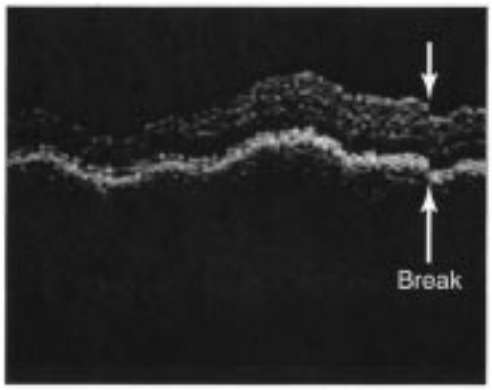

Fig. 11. Another sample OCT scan shows undulations. These larger undulations can result from eye motion or motion of the patient's head. This scan also has a small break.

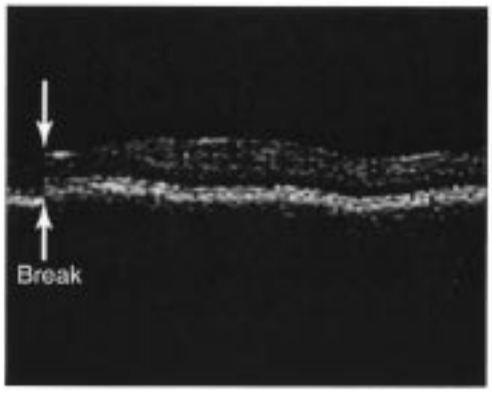

Fig. 12. Another sample OCT scan shows a break in the scan. Scans featuring breaks larger than this would, most likely, not be saved by the OCT technician.

OCT fundamental principles imply that structural interfaces give rise to impedance mismatches that are represented, in turn, by A-scan peaks. However, we instead mark the positively sloped, leading edges of peaks, as they are more consistent and easier to detect. Again, so long as our system consistently detects the same features from image to image, it will be capable of detecting changes in retinal thickness. The detected edges serve as the input primitives to the rest of the system, which then identifies and characterizes the correct boundaries.

To select, organize and interpret the detected edges, we developed a mathematical model for normal boundary contours. We defined a discrete coordinate system (see Fig. 3) where c indexes the A-scans within the image and r indexes the pixel location (retinal depth) within each A-scan. Thus, OCT image coordinates range from $1 \leq c \leq 100$ and $1 \leq r \leq 500$, with 500 at the bottom (outermost side) of the image. While *inner* and *outer* refer to anatomical orientation within the retina, we will use *upper* and *lower* for orientation within the images.

We can model the boundary displacement between adjacent scans as a Mth order Markov sequence. Boundary deviations from one A-scan to the next are due to the actual slope of the retinal surface, patient motion relative to the OCT machine and image noise. Retinal continuity implies that the retinal boundaries in neighboring A-scans should be similar and have similar slope. Moreover, the inertia of the patient's body prevents most patient motion and, thus, most motion-induced boundary displacements from being abrupt. (Rapid eye saccades are a notable exception.) Hence, if $b(n)$ represents the boundary of interest (inner or outer) and n is a particular c-value of interest, then

$$b(n) - b(n-1) = \sum_{i=1}^{M-1} a_i \left(b(n-i) - b(n-i-1)\right) + N(n) \tag{7}$$

where the displacement between $b(n)$ and $b(n-1)$ is a weighted sum of the displacements between adjacent pairs in the previous M A-scans. Note that this is equivalent to representing $b(n)$ as a weighted sum of the neighboring M A-scans, making the boundary model Mth order. Thus, we are using an autoregressive mathematical model for this Markov process. Moreover, $N(n)$ in (7) represents the prediction errors as random noise distributed identically and independently over c. Finally, also note that (7) predicts the boundary locations in an A-scan from the boundary positions to the left of that A-scan. To make predictions using boundary positions to the right of a given A-scan, we use

$$b(n) - b(n+1) = \sum_{i=1}^{i=M-1} a_i \left(b(n+i) - b(n+i+1)\right) + N(n). \tag{8}$$

Note that the two sets of a_i in (7) and (8) are assumed to be direction-specific and, hence, distinct. Thus, we can predict $b(n)$ from the neighboring M boundary locations to either the left or right of n.

We wished to choose optimal a_i based on the retinal boundaries in a large training set. Once the boundary positions for a training set of scans is known, the minimum mean squared error (mse) estimate for the a_i in the Markov model can be found as described by Stark and Woods [24]. We will first build the discussion for (7); the discussion for (8) is very similar. Thus, define the displacement, d, between boundary locations in adjacent A-scans as

$$d(i) = b(i) - b(i-1). \tag{9}$$

Equation (7) predicts $d(n)$ from $d(n-1), \ldots, d(n-M+1)$ and so using the generic labels of $\mathbf{x}$ and y for independent and dependent variables, respectively, and noting that $\mathbf{x}$ is a vector

$$\begin{aligned} \mathbf{x}(\mathbf{n}) &= [\,d(n-1) \quad d(n-2) \quad d(n-3) \quad d(n-4)\,] \\ y(n) &= d(n). \end{aligned} \tag{10}$$

As detailed in Stark and Woods [24], the cross-correlation vector $\mathbf{k_{xy}}$ between $d(n)$ and $\{d(n-1), \ldots, d(n-M+1)\}$ and the autocorrelation matrix, $\mathbf{K_{xx}}$, for

$\{d(n-1), \ldots, d(n-M+1)\}$ are as shown in (11) and (12) at the bottom of the next page.

Here, E indicates that the individual elements in $\mathbf{k_{xy}}$ and $\mathbf{K_{xx}}$ are expected values. Continuing from [24], the minimum mse estimate for the vector $\mathbf{a} = [\mathbf{a_i}]$ is

$$\mathbf{a} = \mathbf{k_{xy}^T K_{xx}^{-1}}. \tag{13}$$

Finally, to predict a_i for (8), we would use similar $\mathbf{k_{xy}}$ and $\mathbf{K_{xx}}$ but redefine the displacement as

$$d(i) = b(i) - b(i+1) \tag{14}$$

and the independent and dependent variables as

$$\begin{aligned}\mathbf{x(n)} &= [\,d(n+1) \quad d(n+2) \quad d(n+3) \quad d(n+4)\,] \\ \mathbf{y(n)} &= \mathbf{d(n)}.\end{aligned} \tag{15}$$

One option for forming the training set to use (13) would have been to manually mark the boundaries in a set of images. However, to avoid this tedium, we selected initial values for a_i and M to start an iterative process that converged to (nearly) optimal values. We started with $M = 5$ and

$$[a_i] = [\,0.2757 \quad 0.1654 \quad 0.0993 \quad 0.0596\,]. \tag{16}$$

Although it might seem startling to see this set of numbers, the rationale is as follows. This initial $[a_i]$ is a weighted sum, where each term is 0.6 times the previous term and the weights were scaled to sum to 0.6; we set these values based on intuition, observation and some trial and error. We then implemented the rest of our algorithm, as described in Section IV below, using this set of a_i for the Markov modeling portion of the algorithm. We applied the algorithm to 330 training images obtained from 14 normal subjects [27], [28]. All images were 3 mm long and all were obtained through the fovea. The initial retinal boundary detection followed the general retinal contours and was acceptable for initializing the training procedure to find the best coefficients for the Markov model.

From each of the training images, then, the displacement between each adjacent A-scan pair was found, according to (9) and (14). We only used the boundary locations from the central 87 A-scans from each image to avoid any possible end effects and we kept the $d(i)$ for the upper and lower boundaries separate. We then grouped the $d(i)$ from each scan into 86 $(\mathbf{x}, \mathbf{y})$ pairs for each boundary, according to (15) for the $[a_i]$ in (7) and according to (16) for the $[a_i]$ in (8). Then, the $(\mathbf{x}, \mathbf{y})$ pairs from all of the 330 training scans were grouped and least squares fits to (7) and (8) were calculated for the upper and lower boundaries.

In particular, let $d_i(n)$ represent the difference between the nth A-scan and its neighbor (left or right depending on context) for the ith OCT image in our training set of 330 images. We will again focus on the development for (7) first. We, thus, define

$$\begin{aligned}\mathbf{x_i(n)} &= [\,d_i(n-1) \quad d_i(n-2) \quad d_i(n-3) \quad d_i(n-4)\,] \\ y_i(n) &= d_i(n).\end{aligned} \tag{17}$$

TABLE I
THE a_i VALUES AS ESTIMATED BY LEAST SQUARES REGRESSION

	a_1	a_2	a_3	a_4
Inner Boundary, Eq. (8)	0.225± 0.005	0.149± 0.005	0.094± 0.005	0.056± 0.005
Inner Boundary, Eq. (9)	0.214± 0.005	0.145± 0.005	0.088± 0.005	0.051± 0.005
Outer Boundary, Eq. (8)	0.240± 0.006	0.107± 0.006	0.051± 0.006	0.022± 0.006
Outer Boundary, Eq. (9)	0.222± 0.006	0.101± 0.006	0.051± 0.006	0.023± 0.006

The data for the least squares fit is then collected from the 330 training images by forming the matrices

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1}(10) \\ \vdots \\ \mathbf{x_1}(94) \\ \vdots \\ \mathbf{x_{330}}(10) \\ \vdots \\ \mathbf{x_{330}}(94) \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_1(10) \\ \vdots \\ y_1(94) \\ \vdots \\ y_{330}(10) \\ \vdots \\ y_{330}(94) \end{bmatrix} \tag{18}$$

and the $[a_i]$ are calculated as

$$\mathbf{a} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T Y}. \tag{19}$$

Note that in (19), the effect of the matrix multiplications and inversions is to create the expected value estimates for (13) and perform the necessary multiplications. Also note that these displacements, $d(n)$, were calculated using the boundaries as found before Step 6, in Section IV. Prior to Step 6, the boundaries represent actual edges in the image; however, Step 6 smooths these boundaries and so destroys some of the information. The least squares fit gave us our second $[a_i]$ and the boundary-detection performance did, in fact, improve over that with the initial $[a_i]$. To further refine the estimate of the $[a_i]$, the boundaries were found using the second set of $[a_i]$ and a new least squares fit was performed to obtain a third and final set of $[a_i]$. Very little difference was detectable between the performance of the second and third $[a_i]$ and so no further iterations were performed. The results from the final least squares fit are in Table I and are in the form mean $\pm$ standard error of the mean. Note that writing (7) and (8) in terms of displacements rather than the actual boundary positions improves the regression. This is because the displacements are typically on the order of 0–10 pixels, while the actual boundary locations, in terms of image coordinates, normally range over 1–500 and they exhibit very little proportional change owing to a sizeable, noninformative constant (average value) term.

$$\mathbf{k_{xy}} = E\,[\,d(n)\,d(n-1) \quad \cdots \quad d(n)\,d(n-M+1)\,]^{\mathbf{T}} \tag{11}$$

$$\mathbf{K_{xx}} = E \begin{bmatrix} d(n-1)\,d(n-1) & \cdots & d(n-1)\,d(n-M+1) \\ \vdots & \ddots & \vdots \\ d(n-M+1)\,d(n-1) & \cdots & d(n-M+1)\,d(n-M+1) \end{bmatrix} \tag{12}$$

TABLE II
INDIVIDUAL STEPS OF OUR ALGORITHM PRESENTED IN A STRUCTURED OUTLINE

Outline of Approach	
1	Input image and filter twice with 4x4 median filter to reduce speckle noise
2	Filter each column with the second derivative of Gaussian, acting as a one dimensional edge detection kernel
3	Choose the strongest two edges, scaled by relative contrast, as the initial approximation to the inner and outer boundaries.
4	Segment the initial approximations at boundary discontinuities
5	Build a continuous boundary from the segments. Start with the longest segment, and first move left and then move right. As boundaries are found, they are considered to be "determined."
5a	Try to link either the initial inner or outer boundary in the new segment with the previous, determined, segment.
5b	Use the Markov model to find the correct inner and/or outer boundary in the new segment if it cannot be linked with those in determined segments.
5c	Check the results of the Markov model for blunders. If blunders are detected, use the original boundaries.
6	Once a continuous boundary is detected, smooth it with a cubic B-spline.

Using the Pearson statistical criterion $p < 0.05$, all of the a_i in Table I are significant and statistically significant differences exist between the a_1 for (7) and (8). This directional asymmetry for our model probably resulted from retinal asymmetries in the 330 training images. For most of these images, the fovea was located on the left-hand side of the OCT scan, which generally caused the retinal thickness to increase from left to right. Similarly, the inner and outer boundary coefficients also appear to be distinct. The decreasing trend in the a_i suggests that the influence on $b(c)$ from its neighbors declines rapidly with distance and that any additional terms will start to model progressively less significant portions of the error. To verify that the optimal model order M is in fact five (resulting, therefore, in four difference values), the performance evaluation as reported below was also performed using models of orders three through nine. Over this range, the performance (not reported here) was very similar, suggesting that this modeling technique is not highly sensitive to order number. However, models of order four and higher exhibited less error than the third-order model, while model orders above showed negligible improvement. A model order of five offers the best compromise between performance and model complexity.

IV. ALGORITHM

We will now walk through the algorithm and we will follow the same OCT image, Fig. 4(B), through the entire process to illustrate the effects of each step. Table II outlines the individual steps in the algorithm for clarification. It is important to note that although the Markov model parameters are established through a statistical evaluation, once set these parameters are kept constant. Thus, for a given B-scan, the algorithm always produces the same result.

A. Median Filtering (Step 1)

We initially applied a 4×4 median filter twice to each image to suppress the speckle noise. The effect on Fig. 4(B) appears in Fig. 13(A) and (B) for the whole image and an individual A-scan, respectively. We see that most of the speckle is removed, while the gross retinal outlines are intact. Though the lack of registration between image columns caused the median filtering to introduce artifacts, the detrimental effects of these artifacts were far less than that of the speckle.

B. Columnwise Edge Detection (Step 2)

For our finite impulse response edge-detection kernel, we used the second derivative of a 1-D Gaussian. If A represents an A-scan and the Gaussian g is defined as

$$g(r) = e^{-r^2/(2\sigma^2)} \tag{20}$$

then

$$s(r) = g'' * A \tag{21}$$

where g'' is the second derivative and $s(r)$ is the output of the edge-detection kernel. The edge-detector scale is determined by σ, the Gaussian standard deviation and the kernel size was set to be 8σ to limit truncation errors [29]. We chose $\sigma = 5$ pixels; the resulting kernel responded strongly to both the inner and outer retinal boundaries in all 330 training images. Smaller values for σ resulted in noisier edges with more clutter, while larger values generated weak responses to the retinal boundaries in some regions. Weak responses typically occurred in the foveal region or in the leftmost or rightmost A-scans. Because the image is metric and because we are always looking for the same structure, a single-scale analysis is appropriate.

This filter is a 1-D version of the Marr–Hildreth operator [26]. As mentioned before, 1-D edge detectors proved to be problematic due to the frequent dislocations between adjacent A-scans. Other 1-D, low-level edge detectors, such as a 1-D Sarkar–Boyer [14], could have been used, provided the algorithm were trained with that detector's responses. These edge detectors' step response is a zero crossing, whose slope and polarity vary with the contrast and polarity of the step[1]. As seen in Fig. 13(B), the retinal boundaries generate narrow peaks in the A-scans, which are confounded by additional peaks from noise and other retinal structures. Each peak generates positive and negative zero crossings when convolved with the kernel, which must all be evaluated to determine which represent the desired retinal boundaries. We developed three criteria to perform this evaluation.

1) Polarity: We first observed that the imaged upper and lower retinal boundaries were always below very dark regions and above bright regions. The upper boundary lies below the vitreous humor and the lower boundary lies below the photoreceptors. Thus, the resulting zero crossings have a negative polarity, which was used as the first criterion for edge validation via

$$s_{\text{NZC}}(r) = \text{negative zero crossing}\,(s)\,. \tag{22}$$

The vector $s_{\text{NZC}}(r)$ equals one where $s(r)$ has a negative zero crossing and zero elsewhere.

[1]A zero crossing response was not required for this problem; an edge detector whose step response is a peak would serve as well.

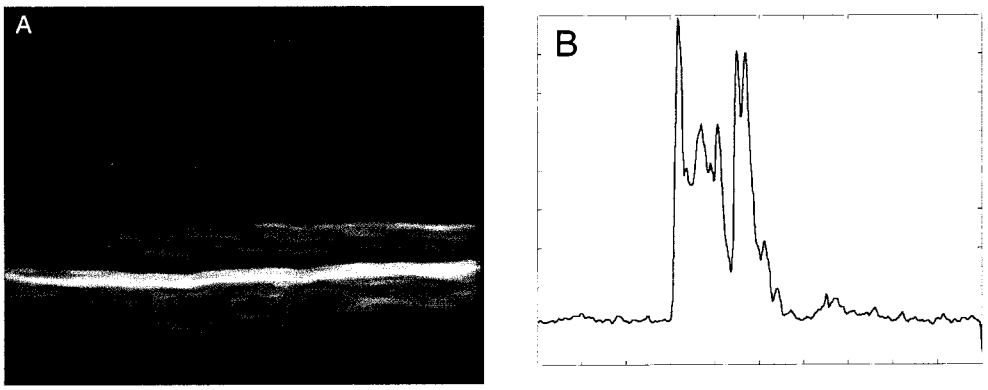

Fig. 13. (A) llustrates a OCT image after median filtering. The original OCT image is in Fig. 4(B). The speckle is greatly reduced and some internal structure is evident within the retina. (B) A-scan after the median filtering operation has been performed on the image. Note that the size of the peaks in the empty portion of the OCT image has been reduced.

2) Slope: The A-scan peaks generated by the retinal boundaries were also characterized as having particularly sharp slopes. We approximated the local slope within the A-scans by convolving them with the first derivative of g and so the *strength* of each edge where $s_{\text{nzc}} = 1$ was defined as $A * g'$. Stronger edges were favored over weaker ones.

3) Relative Strength: Finally, the desired edges within A were unique in having *especially* dark regions above them and so division of the edge strength by the local pixel intensity above each edge[2] emphasized the retinal boundaries. Other retinal features were transitions from bright regions to even brighter regions and were, thus, de-emphasized by this division. The combination of these three criteria defined the *relative edge strength,* s_r, to be

$$s_r(r) = s_{\text{nzc}} \frac{g' * A}{h_r * A}. \tag{23}$$

In (23), multiplication by s_{nzc} selects edges of the proper polarity, while the numerator measures the edge strength. The denominator calculates a weighted average of pixel intensities above each point by convolving A with the causal half of a Gaussian filter, h_r. Here, h_r has a standard deviation 1.5σ; a larger standard deviation was used here for h_r than for g because the dark bands were of relatively larger scale. The relative edge strength, s_r was then used by all subsequent steps in the algorithm. A sample edge map, where the intensity at each point indicates the value of s_r, is presented in Fig. 14.

C. Initial Edge Selection (Step 3)

The upper and lower retinal boundaries in Fig. 14 are, for the most part, the brightest edges in each column; However, they are discontinuous and some spurious edges are brighter. Nonetheless, the initial boundary estimate relies on this general trend by choosing, for each A-scan, the brightest two edges separated by at least ten pixels. We set this minimum separation, based on normal retinal anatomy, to handle cases where one boundary generated multiple, strong edges. The upper edge is assumed initially to be the inner boundary. This simple estimate worked surprisingly well, though it occasionally selected the wrong edges, creating gaps and jumps. Fig. 15 shows the result for image 303. This image is atypical, however, most images had only a few A-scans with dislocated boundaries. Thus, the initial boundary estimate is given reasonable consideration in the final determination of boundary location.

[2]The raw OCT data have a positive dc (constant-valued) offset and are never zero, even in the darkest regions; no problems result from taking this reciprocal.

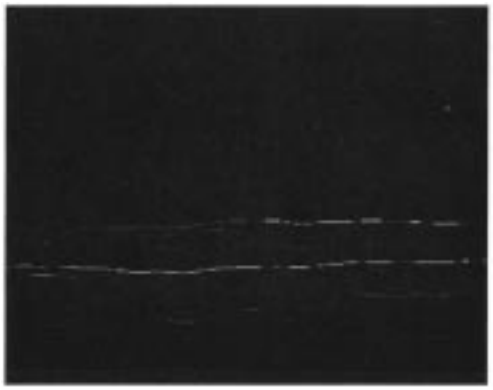

Fig. 14. Edges found by the edge detector in Fig. 4(B). The brightness of each edge is determined by its strength; this is defined by the value of the intensity gradient across the edge, normalized by the image intensity above the edge. Only edges with positive gradient values are shown.

D. Initial Boundary Segmentation (Step 4)

Both initial boundary contours were broken into segments bordered by vertical dislocations larger than ten pixels in *either* contour. Jumps larger than ten pixels are visually displeasing and are likely to be errors in the initial estimate. However, because both contours were segmented together, one contour could be continuous across segment divisions. For the sample image in Fig. 15, the segment divisions are listed in Table III and are illustrated graphically as the vertical lines in Fig. 16. The initial locations for both contours are represented by two dots in each A-scan.

E. Boundary Refinement (Step 5)

We next complete and refine the inner and outer retinal boundaries using information from the Markov model and as much of the initial approximation as possible. Errors in the initial approximation, when present, are usually either *very* large or well within tolerances. Thus, we wish to utilize the instances of small errors and reject the large errors. As a

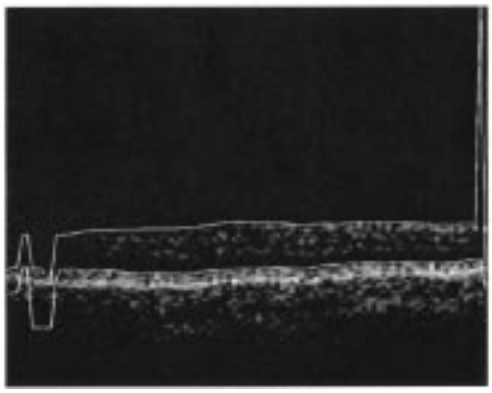

Fig. 15. Results of the initial approximation of choosing the two strongest edges in each A-scan, for Fig. 4(B).

TABLE III
SEGMENTATION OF INITIAL CONTOURS IN THE SAMPLE IMAGE

Segment Number	Segment Length (pixels)	Segment Start (column #)	Segment End (column #)
1	2	1	2
2	4	3	6
3	4	7	10
4	88	11	98
5	1	99	99
6	1	100	100

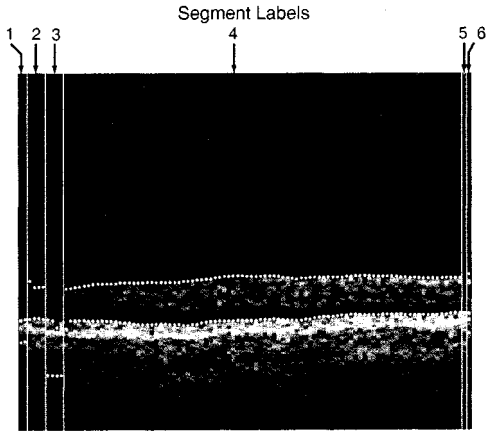

Fig. 16. Segmentation of the initial approximation to the edges for Fig. 4(B). The divisions between the segments are the vertical lines and the boundary location in each A-scan is depicted by a dot. Each segment is labeled by a number corresponding to the segment numbers in Table III.

starting point, we assume that the largest segment (segment 4 in our example) has its inner and outer retinal boundaries correctly identified and located. We then continuously extend these boundaries outward in both directions from this "key" segment, starting with the left direction first. From Table III, the segment order in the sample image is: 4, 3, 2, 1, 5, 6. In this discussion, the boundary segments (either inner or outer) that have already been accepted (the "key" and its extensions) are called *determined* segments; the extension under consideration is called a *new* segment. Also, we remind you that a "discontinuity" refers to a vertical dislocation of ten pixels or more.

1) Connecting the Initial Boundary Approximation (Step 5a): Three distinct cases can occur when extending a boundary between segments and in each case, we determine if either contour in the new segment is continuous with a boundary in the neighboring determined segment. We segmented the initial boundary approximation at discontinuities and so at least one segment contour will always fail the ten-pixel criterion for continuity. The Markov process, as detailed in step 5-2 below, is then used to extend the discontinuous boundaries. The three cases are detailed below:

Case I: The upper contour in the determined segment, now accepted as the inner boundary, has no continuous counterpart in the new segment and the lower contour, now accepted as the outer boundary, has a continuous extension across the segmentation boundary using either the upper or lower contour in the new segment. In this case, we extend the outer boundary using the aligned contour in the new segment; relabeling the contour in the new segment if necessary. See, for example, the extension of the outer boundary from segment 4 to segment 3 in Fig. 16. In this case, segment 3 is the new segment and the initial edge detection has mistakenly chosen choroidal edges for the lower contour and the outer retinal boundary for the upper contour. Thus, we relabel the upper contour in the new segment and join it to the outer retinal boundary in the determined segment. We then use the Markov model, as described below, to find the inner boundary in the new segment as an extension of that in the determined segment. Another example is in the boundary extension from segment 4 to segment 5. These two examples were typical for most common types of discontinuities.

Case II: This is the complement of Case I, with the roles of the inner and outer retinal boundaries reversed. Here we extend the inner boundary using the aligned upper or lower contour in the new segment. The outer boundary will then be extended with the aid of the Markov model.

Case III: Neither line of edge points in the accepted segment is continuous with either one from the new segment. Somewhat surprisingly, this never occurred in any of our 1450 test images. Nevertheless, in this case the algorithm will link neither boundary, but will call on the Markov model to extend both of them through the new segment.

2) Using the Markov Model to Extend the Boundaries (Step 5b): We use the Markov model to improve the selection of edge responses, thereby extending the retinal boundaries beyond the determined segments. Continuing the example for Case I, we must determine the correct inner boundary for segment 3, extending leftward from segment 4. From Table III, we wish to find the boundaries for $7 \leq c \leq 10$ and so using (8) we first predict $b(10)$ as

$$\hat{b}(10) = b(11) + \sum_{i=1}^{4} a_i \left(b\left(10+i\right) - b\left(10+i+1\right)\right) \quad (24)$$

where the sum coefficients are from the second row in Table I. Our algorithm then reevaluates the edges responses in the corresponding A-scan for $c = 10$. Each edge is given a probability score, p, by

$$p(r) = g_p(r) s_r(r) \quad (25)$$

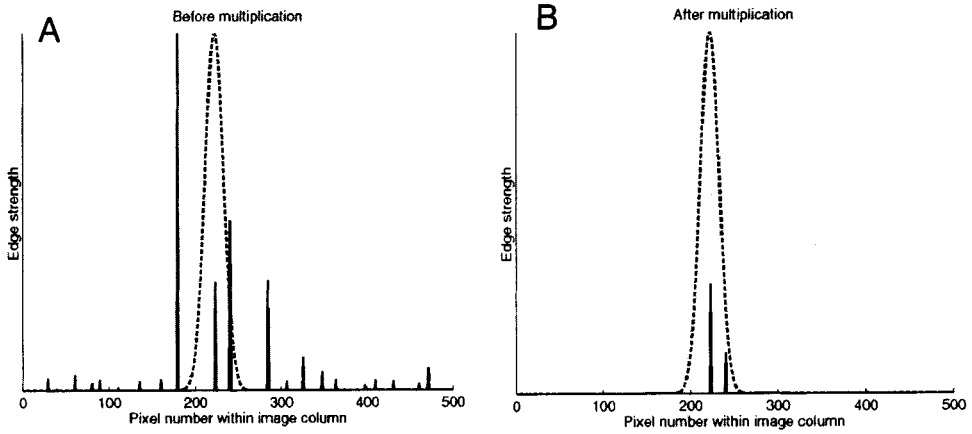

Fig. 17. Sample A-scan shows the Gaussian weighting window centered on the predicted edge location. The boundary we are trying to locate is the outer retinal boundary and the edges are the sharp peaks, where the height of the peak is proportional to the relative edge strength (as defined in (23)). (A) Prior to the multiplication of the edge map by the Gaussian window and (B) after. (B) Edge with the maximum p is then chosen to be the boundary, in this case it turned out to be the one closest to the original prediction.

where

$$g_p(r) = \exp - \frac{\left(r - \hat{b}(i)\right)^2}{2\sigma_p^2} \tag{26}$$

and s_r is from (23). We choose the edge with the largest score and illustrate the outer boundary selection process for a particular column in Fig. 17. Fig. 17(A) is a plot of s_r and g_p; a relatively weak edge is seen to lie close to the predicted location. Fig. 17(B) is a plot of $p(r)$ and we see that the relatively weak edge now has the highest score, while the stronger edges nearby have lower final scores; the highest scoring edge is selected and provides a more continuous boundary extension. Once we determine $b(10)$, we determine $b(9)$ using $b(10)$–$b(14)$ in the same way and we repeat the process up to $b(7)$. Afterwards, the algorithm similarly connects the boundaries in segment 2 and so on. Note that when extending boundaries rightward, as from segment 4 to segment 5, (7) is used. For example, we predict $b(99)$ from $b(98), \ldots, b(94)$ as

$$\hat{b}(99) = b(98) + \sum_{i=1}^{4} a_i \left(b(99-i) - b(99-i-1)\right) \tag{27}$$

where the summation coefficients are now from the first row of values in Table I.

The form of g_p reflects the error distribution about our prediction. We found by trial and error that $\sigma_p = 10$ worked very well. Smaller values for σ_p caused frequent algorithm failures. Seeking an explanation, we found that large, outlying noise errors are not uncommon. In these instances, a narrow Gaussian can force the model to choose a very weak, spurious edge close to the prediction rather than the correct, stronger edge further away. Such errors can then lead the process off the correct boundaries and an example is in Fig. 18. If σ_p is too large, then the predicted $b(i)$ becomes less important as the algorithm reverts to simply choosing the strongest values of s_r.

Checking the Results of the Markov Process (Step 5c): The Markov boundary extension process occasionally makes blunders, which tend to be gradual, progressive deviations from the correct boundary location, without large, telltale dislocations that could be used as markers. By far, he most common error was a merging of the boundary contours, which, once in effect, is propagated by the relative local strength of the retinal boundary edges. Our algorithm checks for this type of error by examining both the final retinal thickness at the far end of the new segment ($b(7)$ for segment 3) and the mean thickness over the segment. *Normal* retinal thickness is never less than 25 pixels (100 μm) and so our algorithm decides an error is likely if either the final thickness or average thickness is less than ten pixels. If so, then any Markov boundary determinations are discarded in favor of the initial boundary estimates, if such a substitution yields a more typical retinal thickness. This rule also accommodates pathological cases where the retina truly is very thin. Obviously, this simple heuristic cannot cope with all possible blunders, but it addresses almost all of those we have observed in extensive testing (1450 images). Continuity between the Markov extension and a contour in the next, new segment was assumed to validate both the Markov extension and the new contour. The new segment's initial contour was then preferred over a continuation of the Markov prediction process into the new segment.

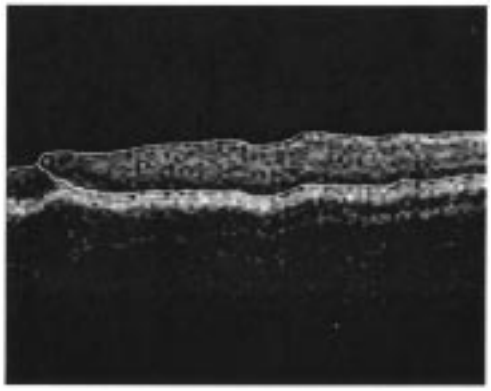

Fig. 18. Example of the type of error that the Markov prediction process can produce. The edge-growing process for the outer boundary has picked the wrong edge near A-scan number ten and this has caused the remaining, further predictions to be wrong, as well. This is the type of error step 5c is intended to detect and correct.

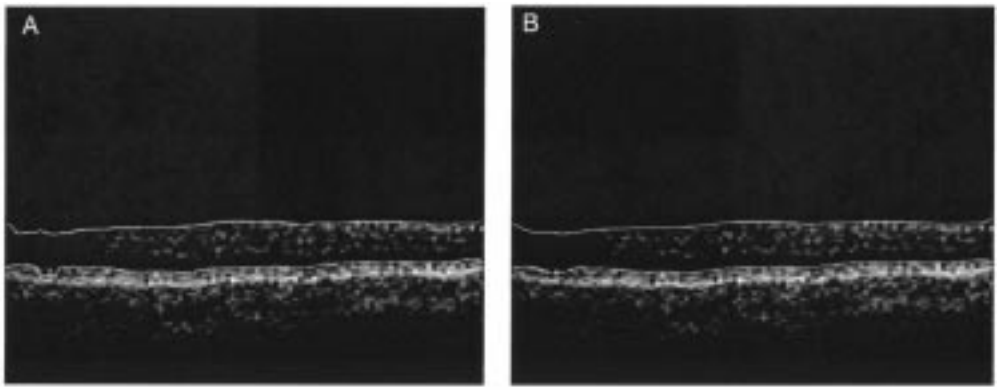

Fig. 19. Final determination of the correct boundary locations for Fig. 4(B). (A) Result after Step 5 in the algorithm, whereby either the original estimate or the Markov prediction is used for the boundary in each segment. Note that there is still a minor discontinuity near A-scan number 10. (B) Result after smoothing by the cubic B-spline.

F. Smoothing the Resulting Boundaries (Step 6)

At this point, we have a set of fairly accurate boundary descriptions of the retina. However, they may still be a bit jagged from minor edge response dislocations. This rough output is shown in Fig. 19(A) Recall that, although we invoke a Markov model of the retinal structure, we use it to select from existing edge responses, not to interpolate between edges or to correct edge positions. We know that the retinal boundary must be smooth, so we now apply a final spline-based adjustment. We compute the centroids of each consecutive (nonoverlapping) group of three boundary points as determined by the procedure given above and fit a cubic b-spline to them. This spline interpolation over all A-scans, an example of which is given in Fig. 19(B), is the final result.

V. Analysis

The clinical application for this algorithm is the measurement of retinal thickness and so we evaluated thickness values resulting from the boundaries rather than directly assessing the boundary placement. We used two different measures of retinal thickness and in each case the algorithm's performance was compared with manually generated ground truth.[3]

We calculated retinal thickness for the nth A-scan, $t(n)$, as

$$t(n) = 4\frac{\mu\text{m}}{\text{pixel}}\left(b_{\text{outer}}(n) - b_{\text{inner}}(n)\right) \tag{28}$$

where b_{inner} and b_{outer} denote the inner and outer boundaries, respectively, and 4μ m/pixel represents the transverse resolution of each pixel. We measured thickness in all 330 training images and, to assess general performance, we used an additional set of 1120 images. To produce ground truth for the thickness measurements, we designed a graphical user interface (GUI) that allowed manual verification and correction of the boundary contours found by the algorithm. Thus, for each B-scan, the thickness measurements from the manually verified boundaries were compared with those using the uncorrected algorithm output and the difference was defined as the *error*. Note that the correction is only for evaluation purposes and is *not* part of the algorithm.

[3]Perhaps more properly termed a "target result."

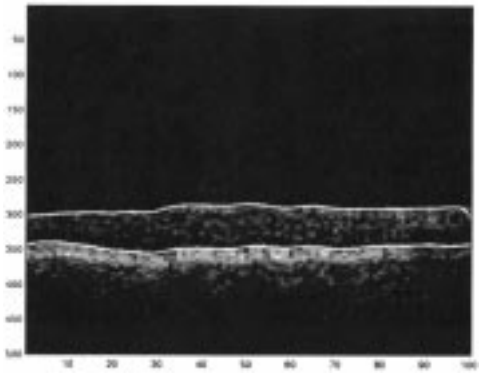

Fig. 20. End effects which can occur. Note how the inner boundary on the right-hand edge of the image drops down through the retina. The boundary determined by the algorithm is the solid line, while the dotted line represents the manual correction.

The corrections were made by one of the authors (DK); representative examples of these corrections are shown in the Results.

We neglected the leftmost and rightmost three A-scans for retinal thickness measurements to avoid end effects induced by the median filtering. An example of these end effects appears in Fig. 20 on the far right-hand side; the solid line denotes the raw algorithm output and the dotted line represents the manual correction.

In clinical practice, one would typically calculate the average retinal thickness over a region of interest (ROI). We defined

$$t_{94} = \frac{1}{94}\sum_{n=3}^{97} t(n) \tag{29}$$

to be the average retinal thickness over the entire scan (a 94-column ROI) and we defined

$$t_{33}(j) = \frac{1}{33}\sum_{n=j-16}^{j+16} t(n) \tag{30}$$

to be the average thickness over all contiguous, 33-column ROIs. Thus, we required $20 \leq j \leq 81$, resulting in 60 different ROIs. For the images we used, 33 columns correspond to a 1-mm-long section of the retina. A physician could conceivably be interested in any portion of a scan and so all 33-column ROIs were considered to detect any weaknesses.

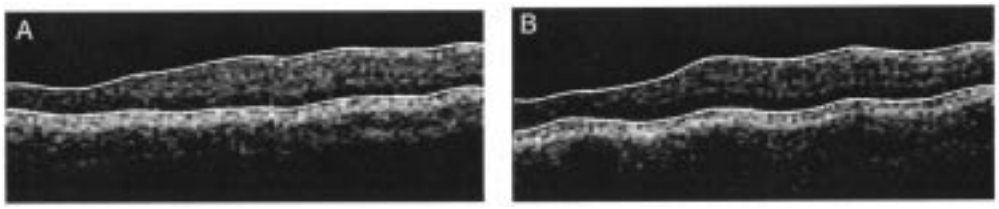

Fig. 21. Boundaries in these images had no (or negligble) error.

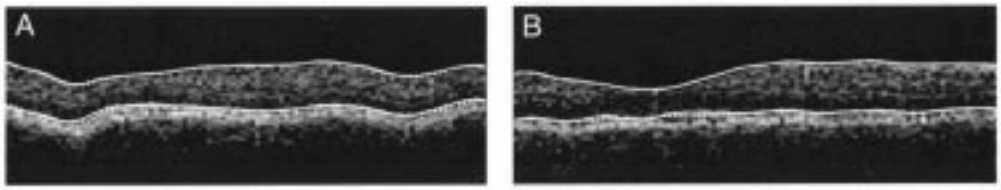

Fig. 22. Insignificant error. These are the same two images as in Fig. 8. (A) Errors are 0.7 pixels for the 94-column ROI and 1.5 pixels for the 33-column ROI; for (B) they are 0.8 pixels for the 94-column ROI and 2.1 pixels for the 33-column ROI.

For the 94-column ROI, the error is

$$e_{94} = t_{94} - t_{94}^{hc} \tag{31}$$

where t_{94} and t_{94}^{hc} are the thickness values from the algorithm boundaries and hand-corrected boundaries, respectively. For the 33-column ROI, this error was defined as

$$e_{33} = \max_j \left| t_{33}(j) - t_{33}^{hc}(j) \right| \tag{32}$$

where t_{33} and t_{33}^{hc} are the thickness values from the algorithm boundaries and the hand-corrected boundaries, respectively. Thus, this measures the worst error among the 33-column ROIs.

We also evaluated the performance of the Humphrey thickness measurements using 466 of the 1450 B-scans. For each of the scans, the Humphrey's thickness values were compared with the same ground truth used to evaluate our algorithm's output and e_{33} and e_{94} were calculated using the same methods described above.

Intersession variability for OCT retinal thickness measurements has been reported to be on the order of 10 μm [27], [30], [31], which is also the OCTs theoretical resolution. Thus, we considered thickness measurement errors less than 10 μm (2.5 pixels) to be *insignificant*. In current clinical practice, thickening is detected by visual assessment using a magnified, stereo view. Therefore, a physician will likely encounter difficulty detecting thickness variations less than 10% of normal retinal thickness. Normal retinal thickness, as measured by OCT, is approximately 255 ± 16 μ m [32] and so 10% is 25 μm (6.25 pixels). Therefore we considered errors between 10 and 25 μm to be *small*. Errors larger than 25 μm were considered to be *large*. We also included an additional performance classification, *failure*. This classification was used in cases where the built-in system located the boundaries completely in the vitreous humor, as in Fig. 8, in cases where the Humphrey system did not find any retinal boundaries and in the one case where our algorithm was grossly wrong, as shown in the Results section.

VI. Results

Qualitative: The boundaries determined by the algorithm generally followed the retinal contours extremely well; the deviations that did occur were most commonly in the outer retinal boundary. Here we illustrate its performance using a cross section of images, as many as space reasonably allows. In the figures, the solid white lines are the algorithm's output, while the dotted lines are the manually corrected boundaries for comparative evaluation. Most scans required very little or no correction, as in Fig. 21. Others suffered noticeable dips or rises in the boundaries, causing larger errors. In Figs. 22 and 23, the error is still small enough to be classified as insignificant. These all come from scans that showed typical errors and, in particular, Fig. 22 presents the same scans for which the built-in algorithm exhibited severe errors. In Fig. 24, the error values qualify as "small." Twenty four of the 1450 images, exemplified by Figs. 25 and 26, contained "large" errors. Fig. 26 is unique in that *it represents the only test in which our algorithm failed to determine the general retinal contours.* Otherwise, the algorithm gave excellent results for the vast majority of the images.

Quantitative: The incidence of each error type is presented in Table IV for the new algorithm. The results for the entire set of 1450 images are listed as well as the results from the set of 330 training images. For the built-in algorithm, the results for the 466 B-scans are in Table V.

VII. Discussion

Our algorithm determines retinal thickness with an error comparable to the 10-μm fundamental OCT resolution and reported intersession repeatability for the vast majority of the 1450 OCT images we tested. Most of the remaining errors were still less than 10% of normal retinal thickness and, thus, represent a substantial improvement over current clinical measurements. There were four basic sources of error.

1) The large amount of speckle noise was not completely removed by median filtering and induced many spurious

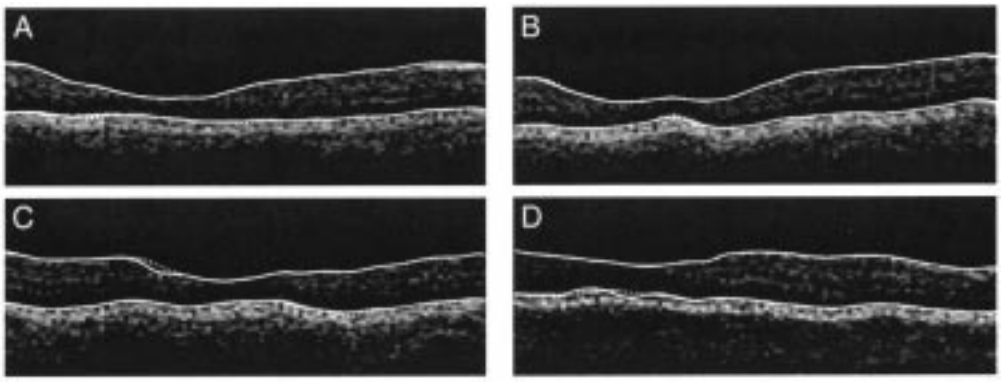

Fig. 23. Insignificant error. For (A)-(D), the errors for the 94-column ROI are, respectively, 0.4, 0.1, 1.1, and 0.1. For the 33-column ROI, the errors are 0.6, 0.6, 2.0, and 1.1. These images are all from Fig. 9.

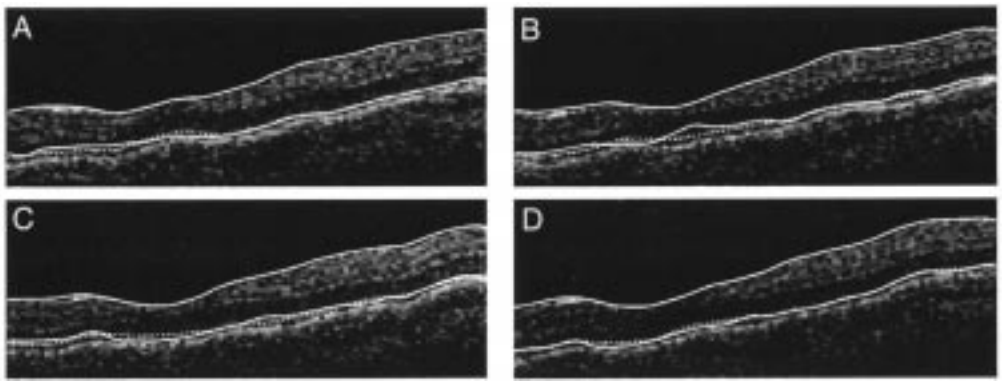

Fig. 24. Small error. For (A)-(D), the errors for the 94-column ROI are, respectively, 0.6, 2.2, 0.2, and 1.6. For the 33-column ROI, the errors are 2.6, 5.1, 3.9, and 3.9. These images are all from Fig. 7.

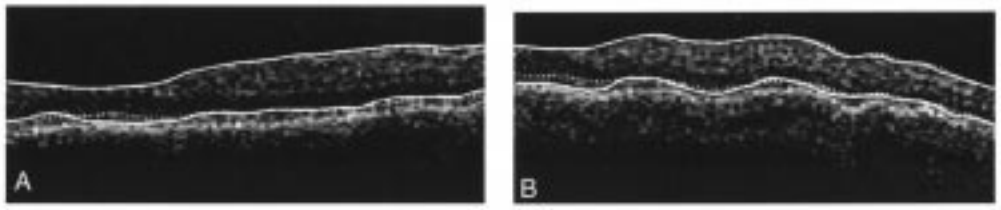

Fig. 25. Large error. (A) Errors are 2.3 pixels for the 94-column ROI and 6.3 pixels for the 33-column ROI. (B) Numerous dips in the outer retinal boundary. The errors are 3.3 pixels for the 94-column ROI and 6.5 pixels for the 33-column ROI.

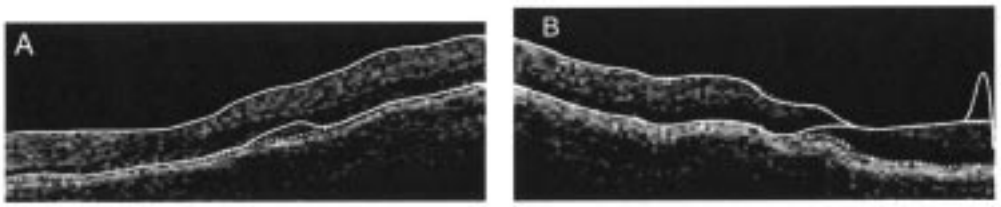

Fig. 26. Large error and failure. These two images are among the four worst errors in our testing group. The image in (A) comes from the image set in Fig. 9. (A) Errors are 3.0 pixels for the 94-column ROI and 27.4 pixels for the 33-column ROI. (B) Very atypical in that our algorithm was unable to determine the gross, retinal contours for this image. The error values are much larger than for any other image and so we consider this a *failure* as in Table IV. The errors are 14.5 pixels for the 94-column ROI and 38.8 pixels for the 33-column ROI.

edges. These edges could occasionally lead the Markov extension process astray, causing errors like those in Fig. 24.

2) The median filtering introduces its own errors as well, by erasing small image features and blurring some of the small discontinuities between adjacent A-scans.

3) Furthermore, the 1-D edge detector selected the leading edges of the retinal boundary peaks. The hand-corrected boundaries were placed relative to the raw OCT images, before median filtering and, moreover, the hand markings tended to be placed closer to the peaks than the edge-detector output Figs. 22 and 23 demonstrate typ-

TABLE IV
SUMMARY OF ERRORS IN DETERMINATION OF AVERAGE RETINAL THICKNESS FOR THE NEW ALGORITHM

	Training Set (330 Scans)		Entire Data Set (1450 Scans)	
Error Type	Whole Scan ROI	33 Column ROI	Whole Scan ROI	33 Column ROI
Insignificant	99.7%	83.0%	97.0%	73.6%
Small	0.3%	16.7%	3.0%	24.8%
Large	0	0.3%	0	1.6%
Failure	0		0.1%	

TABLE V
SUMMARY OF ERRORS IN DETERMINATION OF AVERAGE RETINAL THICKNESS FOR THE HUMPHREY ALGORITHM

	Insignificant error	Small error	Large error	Failure
Whole Scan ROI	83.1%	13.1%	1.0%	2.8%
33 Column ROI	23.0%	65.7%	8.5%	2.8%

ical boundary displacements resulting from these three causes.

4) There was a discrepancy between the retinal boundary model used for the algorithm design and the edges chosen for the hand corrections. Retinal boundaries in OCT images were assumed to be edges featuring large intensity gradients and lying below very dark regions. However, this model can be incorrect, particularly for the outer retinal boundary. The inner choroid can often be seen to have two layers, with the upper layer dimmer than the lower layer; see, for example Figs. 21, 22, 24, and 25. The layered structure is most striking near the fovea, where the two layers can often be seen to separate. There is currently no guidance in the literature as to which layer is the outer retinal boundary. Histological comparisons with OCT images have been done by Chauhan and Marshall [33] and Toth *et al.* [34]. Oddly, however, both of these contributions considered the false color display and not the grayscale images (holding the actual response values) used in this work. The bilayer structure is much less visible in the color images, suppressed by the extremely coarse quantization of the color map. Thus, neither work addressed this issue, nor, in fact, did they look at the fovea. The work of Toth *et al.* was near the fovea, however, and their images suggest that the upper layer corresponds to the outer retinal boundary. In our images, the inner boundary also appears most continuous with the regions away from the fovea and so we chose it for our manual corrections; our algorithm, however, frequently chose the outer layer because of its relative brightness. Furthermore, the delineation between layers was at times highly subjective in noisy OCT images. Thus, in summary, the ground truth for retinal boundary determination has not yet been solidly established.

Our algorithm offers performance significantly superior to the built-in algorithm, the only preexisting system of which we are aware. As can be seen in Tables IV andV, our algorithm offers fewer errors and the errors to which it is prone are less severe. We believe that the sophistication of the Markov model in exploiting spatial coherence lies behind much of the improvement. Even if one only considers the *large* errors and *failures*, the error rate of the new algorithm is significantly lower. Interestingly and in fairness to Humphrey, we point out that the two images in Fig. 26 that caused the worst errors for our system generated only small errors for their algorithm.

Any computer vision system will inevitably produce errors or failures. The GUI we developed, to correct the output of the algorithm during the training phase, would be a useful addition to any clinical OCT system. Boundary correction is quick and efficient and, when used by a trained operator, would give a physician confidence in the measured retinal thickness values.

All images in this study were 3-mm–long horizontal scans through the fovea of normal, healthy retinas. We expect the model to work (almost) without adjustment for different scan lengths through normal retinas, because scans always require 1-s regardless of length and so similar motion artifacts can be expected. Recall that motion artifacts are the dominant mechanism creating the bumps and crenelations.

When extended to the evaluation set, the performance of the algorithm degraded noticeably, but not significantly, from its performance on the training set. This is, of course, to be expected. So long as *insignificant* and *small* errors are satisfactory, the performance is still excellent, with fully 98.3% of the tests yielding errors below clinical significance. This even though the evaluation set contained a much larger mix of images covering a much wider range of image intensities and signal to noise ratios and displaying wider variation in scan placement relative to the fovea.

Measurement accuracy clearly improves with increased ROI size. The boundary placement errors seem random and without bias and so they are reduced when more A-scans are averaged. A 33-column ROI is one third of the scan and is thought to reflect potential clinical use. On another point, the heuristics for detection of errors in the Markov boundary extension algorithm worked very well, exhibiting only one failure in a situation that did not arise in our initial training set. To handle the inevitable errors in any computer vision system, the OCT system should include an onscreen mechanism, such as our GUI, for correcting the algorithm output.

We anticipate that our model will require modification and extension for use with images of retinal pathologies; these images may contain structures that violate our current assumptions. For example, some diseases create fluid filled, retinal cysts that are imaged as dark, empty regions within the

retina. These spaces may present stronger edges than the retinal boundaries. Alternatively, structures within the vitreous humor can generate edges to compete with the inner retinal boundary. More sophisticated techniques, possibly specific to different pathologies, must be developed to distinguish the correct edge responses. The primary problem would be to find reliable "seed" points, as the Markov technique is good at following the correct retinal boundary once it is started. It may also be possible to develop an improved Markov model.

The contribution in this paper can perhaps best be characterized as the first, fundamental building block in a complete OCT retinal analysis system, based on a sophisticated mathematical model of retinal structure. In so doing, this work brings powerful concepts from the area of perceptual organization in computer vision to bear on the problem. For future work, we would like to implement a more sophisticated retinal model that takes the laminar structure of the choroid into account and chooses the correct layer. We would also like to automatically detect vertical deviations in the algorithm's boundaries, but this poses difficult problems. The errors occur in a continuum of sizes and depths, as seen in Figs. 21–26, whose qualitative visual assessment does not correlate well with the numerical error measures. Thus, sharp quantitative cutoffs are impractical. Moreover, it is not clear how to distinguish these vertical deviations from actual pathology causing localized retinal swelling or destruction. Pathological features, like the errors, occur in varying sizes because the OCT lateral resolution varies with the scan length. Ultimately, however, computer vision systems for the analysis of OCT images should be capable of detecting, classifying, quantifying and tracking a range of retinal pathologies based on boundary morphology and retinal image structure.

In conclusion, the retinal boundary-detection system described in this article can determine average retinal thickness to an accuracy comparable to the machine resolution for the vast majority of OCT images. Reliable and accurate measurements of retinal thickness can be expected to improve both the clinical usefulness of the OCT, as well as patient care.

REFERENCES

[1] W. M. Hart, "Adler's physiology of the eye," in *Mosby Year Book*, 9th ed., W. M. Hart, Ed. Chicago, IL, 1992.

[2] F. Ferris and A. Patz, "Macular edema: A complication of diabetic retinopathy," *Survey Ophthalmol.*, vol. 28S, pp. 452–461, 1984.

[3] "Photocoagulation for diabetic macular edema: Early treatment diabetic retinopathy study report," *Int. Ophthalmol. Clin.*, vol. 27, pp. 265–271, 1987.

[4] D. Koozekanani, K. Boyer, and C. Roberts, "Retinal thickness measurements in optical coherence tomography using a Markov boundary model," in *IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 2, Hilton Head, SC, June 13–15, 2000, pp. 363–370.

[5] A. George, J. L. Dillenseger, M. Weber, and A. Pechereau, "Optical coherence tomography image processing," *Investigat. Ophthalmol. Vis. Sci.*, vol. 41, pp. S173–S173, 2000.

[6] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, pp. 321–331, 1988.

[7] F. Lefebvre, G. Berger, and P. Laugier, "Automatic detection of the boundary of the calcaneus from ultrasound parametric images using an active contour model; clinical assessment," *IEEE: Trans. Med. Imag.*, vol. 17, pp. 45–52, Feb. 1998.

[8] J. Wang and X. Li, "A system for segmenting ultrasound images," in *Proc. Int. Conf. Pattern Recognition*, Brisbane, Australia, 1998, pp. 456–461.

[9] A. Sha'ashua and S. Ullman, "Structural saliency: The detection of globally salient structures using a locally connected network," in *Proc. Int. Conf. Computer Vision*, Tampa, FL, December 5–8, 1988, pp. 321–327.

[10] R. Malladi, J. Sethian, and B. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 158–175, Feb. 1995.

[11] R. Malladi and J. Sethian, "A fast level set based algorithm for topology-independent shape modeling," *J. Math. Imag. Vis.*, vol. 6, pp. 269–289, 1996.

[12] K. Lai and R. Chin, "On modeling, extraction, detection and classification of deformable contours from noisy images," *Image Vis. Comput.*, vol. 16, pp. 55–62, 1998.

[13] P. Paillou, "Detecting step edges in noisy SAR images: A new linear operator," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 191–196, Jan. 1997.

[14] S. Sarkar and K. Boyer, "Optimal impulse response zero crossing based edge detectors," *Comput. Vis. Graph. Image Process: Image Understanding*, vol. 54, pp. 224–243, 1991.

[15] R. Aarnink, "Edge detection in prostatic ultrasound images using integrated edge maps," *Ultrasonics*, vol. 36, pp. 635–642, 1998.

[16] A. De Santis and C. Sinisgalli, "A Bayesian approach to edge detection in noisy images," *IEEE Trans. Circuits Syst.-I: Fundamental Theory and Applications*, vol. 46, pp. 686–699, June 1999.

[17] M. Thune, B. Olstad, and N. Thune, "Edge detection in noisy data using finite mixture distribution analysis," *Pattern Recogn.*, vol. 30, pp. 685–699, 1997.

[18] S. Pandit and R. Guo, "Shape mensuration and recognition by DDS approach," in *Proc. 1995 IEEE Int. Conf. Image Processing*, vol. 3, Washington, DC, Oct. 23–26, 1995, pp. 49–52.

[19] M. Das, M. Paulik, Y.-D. Wang, and C. C. Li, "Classification of planar shapes using multiresolution circular autoregressive models," in *Proc. 36th Midwest Symp. Circuits and Systems*, Detroit, MI, Aug. 16–18, 1993, pp. 994–997.

[20] L. Jie and Z. Zhaoying, "Shape recognition using complex nonlinear exponential autoregressive model," in *Proc. 1995 IEEE Instrumentation and Measurement Technology Conf.*, Naltham, MA, Apr. 23–26, 1995, pp. 390–395.

[21] P. Zhou and D. Pycock, "Robust statistical models for cell image interpretation," *Image Vis. Comput.*, vol. 15, pp. 307–316, 1997.

[22] D. Huang, E. A. Swanson, and C. P. Lin *et al.*, "Optical coherence tomography," *Science*, vol. 254, pp. 1178–1181, 1991.

[23] M. D. Kulkarni, C. W. Thomas, and J. A. Izatt, "Image enhancement in optical coherence tomography using deconvolution," *Electron. Lett.*, vol. 33, pp. 1365–1367, 1997.

[24] Stark and Woods, *Probability, Random Processes and Estimation Theory for Engineers*, 2nd ed. Engelwood Cliffs, NJ: Prentice-Hall, 1996.

[25] M. R. Hee, J. A. Izatt, and E. A. Swanson *et al.*, "Optical coherence tomography for ophthalmic imaging," *IEEE Eng. Med. Biol.*, vol. 14, pp. 67–76, 1995.

[26] D. Marr and E. Hildreth, "Theory of edge detection," in *Proc. Roy.Statist. Soc.*, vol. 2076, London, U.K., 1980, pp. 187–217.

[27] D. Koozekanani, C. Roberts, S. Katz, and E. Herderick, "Intersession repeatability of macular thickness measurements with the humphrey 2000 OCT," *Investigat. Ophthalmol. Vis. Sci.*, vol. 41, pp. 1486–1491, 2000.

[28] R. Wang, D. Koozekanani, C. Roberts, and S. Katz, "Reproducibility of retinal thickness measurements using optical coherence tomography," *Investigat. Ophthalmol. Vis. Sci.*, vol. 40, pp. S125–S125, 1999.

[29] G. E. Sotak and K. L. Boyer, "The Laplacian-of-Gaussian kernel: A formal analysis and design procedure for fast, accurate convolution and full frame output," *Comput. Vis. Graph. Image Processing*, vol. 48, pp. 147–189, 1989.

[30] M. R. Hee, C. A. Puliafito, and C. A. Wong *et al.*, "Quantitative assessment of macular edema with optical coherence tomography," *Arch. Ophthalmol.*, vol. 113, pp. 1019–1029, 1995.

[31] M. Baumann, R. C. Gentile, J. M. Liebmann, and R. Ritch, "Reproducibility of retinal thickness measurements in normal eyes using optical coherence tomography," *Ophthalmic Surg. Lasers*, vol. 29, pp. 280–285, 1998.

[32] M. R. Hee, C. A. Puliafito, and J. S. Duker *et al.*, "Topography of diabetic macular edema with optical coherence tomography," *Ophthalmology*, vol. 105, pp. 360–370, 1998.

[33] D. S. Chauhan and J. Marshall, "The interpretation of optical coherence tomography images of the retina," *Investigat. Ophthalmol. Vis. Sci.*, vol. 40, pp. 2332–2342, 1999.

[34] C. A. Toth *et al.*, "A comparison of retinal morphology viewed by optical coherence tomography and by light microscopy," *Arch. Ophthalmol.*, vol. 115, pp. 1425–1428, 1997.