Repeatability Characterization and Computer Vision Based Analysis of Optical Coherence Tomography

DISSErTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Dara Koozekanani, MS BME, The Ohio State University

* * * * *

The Ohio State University

2001

Dissertation Committee:

Cynthia Roberts, PhD, Co-Adviser
Kim Boyer, PhD, Co-Adviser
Steven Katz, MD

Approved by

Co-Adviser

Biomedical Engineering Program
ABSTRACT

Optical coherence tomography is a new imaging technology providing high resolution, cross sectional views of the retina. These images detail microscopic retinal pathology, and are particularly promising for their ability to yield objective, numerical measurements of retinal features in vivo. These measurements may help improve treatment protocols for retinal pathologies involving edema or destruction of tissue layers. However, the measurement repeatability has not been well characterized in the literature. Moreover, there is not a good system available for automatically making thickness measurements from images. Finally, subject motion during scan acquisition causes uncertainty in the precise retinal region imaged. In this dissertation, we address all of these shortcomings, using clinical studies to assess the repeatability and computer vision to present solutions to the other two problems.
To my grandmother, father, and mother
ACKNOWLEDGMENTS

I would like to thank Ed Herderick, Ashraf Mahmoud, Rick Campbell, and Vladimir Marukhlenko for all their help with my research projects.
VITA

1994 ......................................................... B.S. Electrical Engineering
2000 ........................................................ M.S. Biomedical Engineering
1994-present  ........................................... MD/PhD student
The Ohio State University

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Biomedical Applications of Computer Vision

Studies in:

- Computer Vision: Prof. Kim Boyer
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xxi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Statement of The Problem</td>
<td>1</td>
</tr>
<tr>
<td>2. Significance</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Ophthalmic Significance of Thickness Measurements</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1 Diabetic Macular Edema</td>
<td>5</td>
</tr>
<tr>
<td>2.1.2 Glaucoma</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Thickness Measurement Repeatability</td>
<td>5</td>
</tr>
<tr>
<td>2.2.1 Within-session Repeatability</td>
<td>6</td>
</tr>
<tr>
<td>2.2.2 Intersession and Long Term Repeatability</td>
<td>7</td>
</tr>
<tr>
<td>2.2.3 Repeatability with Power Range and Dilation State</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Automatic Thickness Measurement</td>
<td>10</td>
</tr>
<tr>
<td>2.4 Video Tracking of Patient Eye Motion</td>
<td>11</td>
</tr>
</tbody>
</table>
10. Conclusions ................................................................. 220
    10.1 Current Achievements ........................................... 220
    10.2 Limitations ............................................................. 221
    10.3 Future Directions .................................................. 222

Bibliography ........................................................................... 223
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 A-D. These images show how 2 B-scans acquired from the same subject but under different scanning protocols can appear very different, regardless of the display method. Parts A and B show the two scans as they would be displayed by the Humphrey 2000 during acquisition, while parts C and D show the same two B-scans using normalized, grayscale color maps. The higher intensity B-scan (A and C) was acquired with dilated pupils and a 750 W SLD power setting, while the lower intensity one (B and D) was acquired through undilated pupils and a 500 W SLD power setting. As described in a later chapter, the average scan intensities for the high and low intensity B-scans are 1075 and 422 intensity units, respectively, and their measured retinal thickness values are 275.0 ( \mu m ) and 281.3 ( \mu m ) respectively.</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Here is an example of pupillary shadowing of the OCT image.</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Here we have some samples of erroneous boundary detections by the Humphrey system. In the left image, note the boundary being displayed up in the vitreous humor, while on the right, the detected boundary lies well within the retina.</td>
<td>11</td>
</tr>
<tr>
<td>3.1 A schematic drawing of the OCT is shown here, emphasizing how it is essentially a Michelson interferometer. The outgoing light paths are solid lines, while reflected light is drawn as dashed lines.</td>
<td>15</td>
</tr>
<tr>
<td>3.2 A 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.</td>
<td>17</td>
</tr>
<tr>
<td>3.3 A sample OCT scan from the set of training images, with anatomical features labeled.</td>
<td>18</td>
</tr>
</tbody>
</table>
3.4 Yet another OCT scan is presented here to further demonstrate the possible variations. Figures 3.4A and 3.4B are different scans from the same person, while Figure 3.3 is from a different person. Note that in Figures 3.4A and 3.4B, 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.

3.5 Here we have a sample A-scan from Figure 3.4A, without any filtering. The DC offset is evident, as is the noise. The inner and outer retinal boundaries have been labeled.

3.6 This image displays artifacts (indicated by arrows) that can result from shadowing of the retina by the pupil margins. The result is that the retina is obscured in the shadowed region. This image is displayed by the Humphrey 2000 using a normalized color map.

3.7 This figure shows another example of how two B-scans, acquired from the same subject, can appear very different regardless of the color map used for display. The differences persist even with the normalized color maps, indicating that they are not simply a result of scaling. The high intensity B-scan is on the left, and the low intensity B-scan is on the right. Parts A and B use the Humphrey absolute color map, C and D use the Humphrey normalized color map, and E and F use a normalized grayscale color map. As described later in the paper, the average scan intensities for the high and low intensity B-scans are 1023 and 652 intensity units, respectively, and their measured thickness values are 283.3 μm and 277.6 μm respectively.

3.8 This figure shows more examples of acceptable performance of the Humphrey boundary detection system.

3.9 This figure shows examples where the Humphrey boundary detection system completely failed.

3.10 This figure shows examples of errors that typically occur with the Humphrey boundary detection system. These errors are large enough to be objectionable and disrupt measurements.

3.11 Here is a sample screen capture for the Humphrey system, showing the information that is given to the user.
4.1 This figure shows a fundus view of the OCT scan position, as shown by the white line. The white haze above and below the line is a glare artifact. .............................................. 52

4.2 This figure shows a sample OCT scan with the measurement boundaries marked. ............................................................................................ 53

4.3 This figure shows the retinal thickness contours from five scans obtained within a single session. ................................................................. 54

5.1 A sample fundus view is shown with the scan placement illustrated as a bright, horizontal line located nasally to the fovea. .................. 65

5.2 A typical OCT scan is displayed in grayscale, with various anatomical features marked ............................................................................. 65

5.3 This sample OCT scan has the retinal boundaries, as found with the custom software, outlined in white, and the regions measured for the thickness analysis have been shaded with gray rectangles. In the figure, the black vertical line denotes the foveal center. .............................. 66

5.4 A and B: These figures illustrate, for two different subjects, the overall similarity of the retinal thickness curves generated from the two different visits. In each figure, the dashed lines denote the five scans obtained at the first session of the initial visit, and the solid lines denote the five scans obtained during the second visit. The individual plot lines have been shifted horizontally to align the foveal center. . . 69

6.1 This sample OCT scan has the retinal boundaries, as found with the custom software, outlined in white, and the regions measured for the thickness analysis have been shaded with gray rectangles. In the figure, the black vertical line denotes the foveal center. ......................... 75
6.2 These scatter plots show that varying the scanning parameters has no consistent effect on the thickness measurements (fovea in A and 1 mm retinal region in B), as illustrated by the fact that the stars in each column are all centered about zero. The four columns in each plot represent one set of scanning parameters, and each star represents the difference between a subject's measurement under that set of parameters and their average measurement over all sets of parameters. The four sessions are labeled on the horizontal axis; "u500" denotes the Undilated-500 μW session, and likewise for the other three sessions.

6.3 These scatter plots demonstrate the lack of effect of B-scan signal intensity on measured retinal thickness. The axes are for relative values of intensity and thickness to remove the effect of between-subject thickness differences.

7.1 A sample OCT scan showing the undulations which may be present. These undulations may be the result of tiny, involuntary eye motions.

7.2 Another sample OCT scan showing undulations. These larger undulations can result from eye motion or motion of the patient's head. This scan has a small break near column 90.

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

7.4 This figure illustrates a OCT image after median filtering. The original OCT image is in Fig. (3.4B). The speckle is greatly reduced, and some internal structure is evident within the retina.

7.5 This figure shows an 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.

7.6 This figure shows the edges found by the edge detector in 3.4B. 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 depicted here.
7.7 This figure shows the results of the initial approximation of choosing the two strongest edges in each A-scan, for Fig. (3.4B). .......................... 100

7.8 This figure shows the segmentation of the initial approximation to the edges for Fig. (3.4B). The divisions between the segments are the vertical lines, and the boundary location in each A-scan is depicted by a dot. Above the image, each segment is labeled by a number corresponding to the segment numbers in Table 7.3 .......................... 101

7.9 Here is a sample A-scan from Fig. (7.9) showing the Gaussian weighting window centered on the predicted edge location. The boundary we are trying to locate is the outer retinal boundary. The edges are the sharp peaks, where the height of the peak is proportional to the relative edge strength (as defined in Eq. (7.17)). .............................. 105

7.10 This figure depicts the A-scan in Fig. (7.9) after the Gaussian window is multiplied by the edge map. The 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. .............................. 106

7.11 Here is an 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 10, 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. .............................. 107

7.12 This figure shows the final determination of the correct boundary locations for Fig. (3.4B). This is the 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. .............................. 108

7.13 Here are the final boundaries after the intermediate results in Fig. (7.12) have been smoothed using the B-spline. .............................. 109

7.14 This image shows the 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. .............................. 110

7.15 The boundaries in these images had no error. .............................. 114
7.16 Insignificant error. These are the same two images in Fig. (3.9). In part A, the errors are 0.7 pixels for the 94 column ROI and 1.5 pixels for the 33 column ROI; for part B they are 0.8 pixels for the 94 column ROI and 2.1 pixels for the 33 column ROI.

7.17 Insignificant error. For part 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. (3.10).

7.18 Small error. For part 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. (3.8).

7.19 Large error. In Part A, the errors are 2.3 pixels for the 94 column ROI and 6.3 pixels for the 33 column ROI. Part B features 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.

7.20 Large error. These two images are among the four worst errors in our testing group. The image in Part A comes from the image set in Fig. (3.10). Part B is 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. The errors are -14.5 pixels for the 94 column ROI and 38.8 pixels for the 33 column ROI.

8.1 A GUI for adjusting retinal boundaries. The controls are the buttons at the bottom.

9.1 Here are sample video fields as captured from two different subjects using the Humphrey video camera. Note the low contrast and poor focus.

9.2 This series of video fields, obtained in sequence, demonstrate how the image can suddenly blur. The middle image in particular shows almost no features.

9.3 Here is one sample video frame from each of the six training subjects used in this study.
9.4 Here is a scan image showing the circular path a scan may take around the optic nerve head; the path is illustrated as the bright circle. This view was created by the Humphrey system after the scan was acquired.

9.5 Here are some sample fields showing variations that can occur in the size and number of SLD spots.

9.6 These images show the appearance of the blood vessels in the video images. The left image shows the entire video field, with a smaller region selected in the box. The middle image shows this box close up, while the right image shows a single, representative column from the middle image.

9.7 The vessel distribution in these images is a pattern for our idealized, radial distribution of blood vessels. Note how many large vessels appear to come out from the nerve head center. Note too that many vessels don’t follow this pattern.

9.8 Here are some sample training subimages used to train the nerve head eigenspace. Shown here are two sample training images from each subject.

9.9 Here are some sample training images used to train the retinal eigenspace. Again, we have two sample images from each subject.

9.10 Here we see how the eigenspace error decreases as we increase its dimensionality, up to a point.

9.11 Here are the eigenvectors representing \( \hat{\Phi}_N \).

9.12 Here are the eigenvectors representing \( \hat{\Phi}_R \).

9.13 This image shows how the dual eigenspace approach is able to localize the nerve head in a video image. On this and the next few pages, we show two sample fields for each subject. We start with the video field in the upper left, followed by \( \pi_N \) and \( \pi_R \). The bottom row show \( \eta \), in intensity form and mesh form for emphasis. The video images are shown with a reverse color map, so that darker implies larger values of \( \pi_N \), \( \pi_R \), and \( \eta \).

9.14 Here are more samples.
9.15 Here are more samples................................. 150
9.16 Here are more samples................................. 151
9.17 Here are more samples................................. 152
9.18 Here are more samples................................. 153
9.19 Here are some sample fields showing the effect of the minimum filter (pre-filter is on left, post-filter is on right) ...................... 157
9.20 Here are some sample fields showing the clusters that were organized into bright spots by the SLD finder subsystem. .................... 159
9.21 Here are some sample fields showing the labels given to spots in different video fields. .......................................................... 161
9.22 Here is a matrix indicating the video fields in which each labeled spot existed; white means existence, black means not existing. The bottom row is the landmark spot; note the regular pattern. The top row is the scanning spot; note that it is there for most fields (the middle row is glare). .................................................. 163
9.23 Here are some sample fields showing how the SLD spot can be masked off. In each column, the top image is without the masking, the middle image is using the bilinear transform without the addition of noise, and the bottom image is with the addition of noise. ...................... 167
9.24 These are the 1-dimensional filters used to detect blood vessels in the retinal image .......................................................... 169
9.25 Here are the sample step and pulse responses of the 1-D row filter; the responses of the column filter are similar. The pulse used is a 10-unit long, positive pulse; this pulse simulates the SLD spot in the image. .. 170
9.26 Here are some sample fields showing the result of the 1-D vessel filter. There are two video fields from 4 subjects and one from a 5th, and, in each case, we have a row of video fields above a row of filter results. 172
9.27 Here are some more fields showing the result of the vessel filter. . . . 173
9.28 Here is a sample figure showing how the SLD spot can dominate the filter output when it is present. This is why we mask it off. A masked filter result is on the left, and an unmasked filter result is in the middle. One can see that the region of the laser spot can be very negative compared to the rest of the image; a mesh plot of the middle image is on the right so that the relative magnitude of the spot region’s response can be appreciated.

9.29 This shows how thresholding the video images preserves the vascular structure while mitigating the noise. The middle image is the intensity thresholded image, \( I_I \). The rightmost column shows the size thresholded image, \( I_J \).

9.30 This shows additional examples of intensity and size thresholding.

9.31 This shows the result of peak detection in \( I_I \) (left column) and \( I_I \), yielding \( I_P \) (middle column) and \( I_P \) (rightmost column).

9.32 Here are more examples.

9.33 This shows the directional templates used in the organizer. On the left are the templates used to vote for new growth directions. On the right are the templates used to zero out points in \( I_P \) after voting has occurred. In these binary template images, black equals 1 and white equals 0.

9.34 The leftmost image in each row shows \( I_I \) so that we can show the structure that is there to be found. The middle image shows the binary image of detected vessel positions, \( I_v \). The right image shows the positions of the detected contours, \( \xi_j \) overlaid on top of the original video field for comparison.

9.35 Here are more examples.

9.36 This shows the templates used in the Hough transform. To the left is the template used for the \( \rho_h = 17 \) Hough space, and to the right are the templates used for \( \rho_h = 33 \) Hough space. In the bottom right of each figure is the template used when no directional information was available; this template is a full circle.
9.37 This shows samples of $H$ for the $\rho_h = 17$ Hough space on the left and the $\rho_h = 33$ Hough space on the right. In the top row, the directional information is used, and in the bottom row it is not. Though the difference is not dramatic, one can see that the top row is somewhat less cluttered, with more contrast between the peak regions and their surroundings.

9.38 This shows samples of $H$ for the $\rho_h = 17$ Hough space on the left and the $\rho_h = 33$ Hough space on the right. In the top row, an intensity threshold has been applied so that only the points greater than 2.5 standard deviations are used, and in the bottom row, the top 100 of these points are chosen.

9.39 This shows all the circles found by the Hough transform. The left column shows $I_p^1$ to illustrate the available structure. The middle column shows the detected circles overlaid on $I_p^1$, while the right shows them on the video field.

9.40 Here are more examples.

9.41 This shows how the eigenimage technique chose circles. The left column shows $I$ with all detected circles superimposed. The middle shows $\eta$ calculated for the entire video field, while the right shows the chosen five circles.

9.42 Here are more examples.

9.43 Here are some samples of nonlinear contours, $\xi_j$ and their relations to the circles $(r_i, c_i, \rho_i)$. On the left is a contour that doesn't touch the circle, in the middle is one that touches, and on the right the contour is tangential to the circle.

9.44 Here we see samples of $\xi_j$ that are along the perimeter of the circle and hence circumferential.

9.45 Here are some samples of linear contours that are aimed to the center of the circle.

9.46 Here are some samples of linear contours that are not aimed to the center of the circle.
9.47 Here are some exceptions to the model. On the left is an example of a linear contour that is aimed to the center of a spurious circle. 206

9.48 Here are samples showing that $\xi_j$ intruding into a circle was usually, but not always, an indication that the circle was not the nerve head. 206

9.49 Here are cases where the long contours corresponding to vessels have generated spurious responses in the Hough space that are tangential to the vessels. 207

9.50 Here are a pair of images showing how the correct circle is often closer to the ideal, uniform spoke distribution of vessels. The rightmost image is closer to the ideal than the leftmost; note though that for the leftmost, though the detected vessels are oriented towards the center of the circle, they do not enter through a wide range of angles. 208

9.51 Here are examples of how $V$ converges to a coherent model of the vascular distribution. The left images show sample video fields to illustrate the actual vasculature. The middle images show $V$ after 10 video fields, and the rightmost shows $V$ after 50 fields. 211

9.52 Here are more examples 212

9.53 Here are samples of $C$, the correlation between $I_o$ and $V$, to illustrate how sharp the peak usually is. 214

9.54 Here are examples of our final results. In each case, the images are grouped in sets of two rows. In the upper row, the positions of the nerve head through the sequence of video fields is shown; a dot is placed at the nerve head's position in each video field. The positions are superimposed on the first video field of the sequence. In the lower row, we see the actual path the SLD beam travelled, displayed as a white line. On the line are gray dots indicating precisely where the SLD beam was actually detected. All positions are relative to the initial video field, which is displayed as the background. The path the SLD beam took in the video field is shown as a black circle. 218
LIST OF TABLES

Table | Page
--- | ---
4.1 This table shows the short-term, between-session repeatability for OCT measurements, by showing the means for the various sessions. Retinal thickness measurements are grouped by session (5 scans per session), showing that the thickness means do not vary between the individual sessions. | 57
4.2 This table shows the lack of a within-session effect for OCT measurements, by showing that there is no systematic change in thickness value as one repeatedly measures within a session. In the table, retinal thickness measurements are grouped by session (5 scans per session). | 58
4.3 This table shows that single measurements tend to cluster close to an individual subject's thickness mean, demonstrating that the measurements are highly repeatable. When session averages are considered, the variability is even less. In the table, the differences between subject thickness averages and thickness measurements from individual scans and sessions are shown. | 58
5.1 This table illustrates the similarity of the long and short term variability measures by presenting the intersession and within-session standard deviation (SD) and coefficient of variation (CV) for the two retinal thickness measurements used in this study. Because the three sessions span two visits separated by an average of 15 months, the intersession values reflect long term variability while the within-session values reflect short term variability. | 70
6.1 Reported above are the average, between-session differences in thickness measurements for this study. The session conducted with dilated pupils and 750 $\mu W$ SLD power setting has been arbitrarily taken as a standard to compare sessions in which the scanning power is varied or the pupils are undilated. | 78
7.1 This table has the values for the boundary detector's Markov model coefficients. The $a_i$ are shown as estimated by our least squares regression 92

7.2 These are the individual steps of our boundary detection algorithm presented in a structured outline ................................. 94

7.3 This table shows the segmentation of initial boundary contours of the sample image we are using to illustrate the boundary detection system 101

7.4 This table summarizes the errors produced by our boundary detection system in determining the average retinal thickness of our test images 113

9.1 The error performance ........................................... 216
CHAPTER 1

STATEMENT OF THE PROBLEM

Optical Coherence Tomography (OCT) is a new imaging modality providing cross-sectional images through tissues that are qualitatively similar to ultrasound images. However, the use of near-visible infrared light allows OCT images to have a far higher resolution than ultrasound. Also, the use of light allows OCT to image structures within the eye which are too deep within the body for high frequency ultrasound to image but which can be seen through the transparent vitreous humor. Two applications of OCT that have appeared in the literature [1, 2, 3] are thickness measurements of both the retina and the retinal nerve fiber layer (RNFL). The main contribution of OCT over other existing technologies is its ability to make quantitatively precise measurements that can be compared over time. The Humphrey OCT system, which to the best of the author's knowledge is the only commercially available OCT system in the United States, comes from the manufacturer configured to measure both of these thickness values. Currently, however, the Humphrey, as a retinal thickness measurement system, suffers from three main deficiencies: (1) the lack of published information regarding its variability; (2) the lack of a reliable, automatic means to measure retinal thickness using scan images; and (3) the lack of a means to ensure eye motion does not corrupt measurement information.

1
First of all, thickness measurements made with OCT have not been fully characterized in the literature for any OCT system, much less for the Humphrey 2000. In particular, the variability of its measurements is unknown. There are several aspects of variability, all of which must be investigated: (1) within a single session, (2) across different sessions, (3) over long term separations, (4) under different power settings, and (5) under different dilation states.

To facilitate the acquisition of thickness measurements, clinicians need a fast, reliable, and automatic system to derive measurements from OCT images. Simply stated, they don't have time to make manual measurements from every image. Such a system would also facilitate our research into the measurement repeatability. The software that comes with the Humphrey 2000 OCT system is neither accurate nor reliable. Thus we also developed a system to automatically measure retinal thickness from OCT images. For the repeatability studies, we needed a reliable and accurate system because we are interested in the fundamental repeatability of the Humphrey 2000 OCT system's hardware, and we do not wish to introduce additional errors from the image analysis. We will only investigate thickness measurements of the whole retina. Thickness measurements depend on boundary determination, and the inner and outer retinal boundaries are far more objective and easier to manually distinguish in OCT B-scan images than are the retinal nerve fiber layer (RNFL) boundaries. Thus an algorithm for whole thickness measurement can be manually verified. RNFL measurement, on the other hand, is an empirical determination of the thickness between the inner retinal boundary and the location within the retina where the B-scan signal intensity is decreased by a fixed number of dB. Thus it would
be very difficult to verify, and, moreover, the heuristic for its measurement relies on the correct determination of the inner and outer retinal boundaries.

Finally, we are also concerned with the detrimental effects of eye motion during OCT scan acquisition. The Humphrey OCT acquires B-scan images over a 1.5 second interval, during which both voluntary and involuntary eye motions may occur. Such motions may not be detected by the OCT scan operator, and consequently may adversely affect thickness measurement and increase variability. To alleviate this problem, we wish to track the scan position on the retina. We will track the scan path using digitized video frames from the live video image obtained by the Humphrey 2000 OCT system during scan acquisition. Our focus will be on tracking eye motion during circular OCT scans about the optic nerve head. One major reason for this will be tractability and our desire to be the first to test the concept of tracking OCT scans. In order to track the scan position relative to the retina, retinal landmarks must be found as reference points. However, because of the poor quality of the Humphrey video, the only retinal region with a large, reliable number of features is about the nerve head. If video tracking is to work anywhere on the retina, it must work on the nerve head; if this can be shown to be feasible, then future work may extend it to tracking over the entire retina.

Thus, we have three primary goals for this research. First, we wish to characterize the fundamental repeatability of thickness measurements made with the Humphrey 2000; secondly, we wish to develop a robust and accurate thickness measurement system; and finally, we wish to develop a system to track scan location during eye motion.
CHAPTER 2

SIGNIFICANCE

2.1 Ophthalmic Significance of Thickness Measurements

Measurement of retinal thickness is important for the treatment and diagnosis of many diseases. Current clinical practice is to visually estimate retinal thickness by viewing the retina with a slit lamp biomicroscope, which offers a highly magnified, binocular view of the retina as illuminated obliquely by a narrow slit beam. Both edematous and atrophic retinas have qualitatively different appearances that allow them to be distinguished from normal retinas, irrespective of apparent thickness. However, the numerical thickness of the retina is clinically important, and the lack of precise measurement techniques limits the clinician’s ability to assess a patient’s state. This limitation is particularly evident when one is trying to detect whether subtle changes have occurred between office visits. These differences, though subtle, may determine the particular path a patient’s treatment takes. In contrast with slit lamp biomicroscopy, OCT gives objective, numerical measurements of retinal thickness. Two particular diseases that could benefit from precise retinal measurements are the edema caused by diabetes and the atrophy caused by glaucoma.
2.1.1 Diabetic Macular Edema

Macular edema is the most common cause of vision loss in diabetics [4]. For treatment purposes, macular edema is classified into clinically significant macular edema (CSME) and non-CSME, based on its morphological characteristics. It is important to distinguish these because only CSME has been shown to benefit from treatment [5]. The ability of OCT to give numerical measurements of retinal thickness may be of use in treating CSME because it may allow ophthalmologists to precisely characterize and monitor a patient's edema. Also, with the knowledge that the measurements are highly repeatable, the measurements may be used to derive new treatment standards.

2.1.2 Glaucoma

Glaucoma is the leading cause of preventable blindness in the United States [6]. Two million Americans have glaucoma, with 80,000 of these being blinded from the disease. One of the earliest detectable signs is destruction of the retinal nerve fiber layer (RNFL). The destruction occurs long before visual field loss is evident, possibly because the redundancy in human retinas allows some destruction without loss of function. This destruction can be localized or diffuse, and most glaucoma patients feature both [7]. OCT offers the potential to directly measure RNFL thickness, which may allow more sensitive diagnoses and monitoring of treatment.

2.2 Thickness Measurement Repeatability

Characterization of thickness measurement repeatability is crucial for clinical OCT applications; the normal variability must be known to construct proper confidence intervals around patient measurements. Knowledge of the repeatability is important
both for screening purposes and for interpreting thickness changes between visits. The current lack of characterization is likely impeding the general acceptance of the Humphrey 2000 OCT system in the ophthalmic community at large. Ideally, variability needs to be known under five conditions: within session, between recent sessions, between distant sessions, with variations in OCT scanning power, and with variations in subject dilation.

2.2.1 Within-session Repeatability

This is the variability within a series of scans obtained in a single sitting, when the patient has been positioned and has not knowingly moved. This variability is caused by the inherent physical principles of OCT, instabilities in its hardware, and baseline patient motion. Studies in the literature have reported within-session repeatability \[1, 8, 2\], but they have all used machines that are not commercially available. The prototype OCT systems, which, to date, have dominated the literature, demonstrate significant differences with the commercially available system used in this study. Differences exist in the optics, internal amplification of the interference signal, output interface, and superluminescent diode (SLD) power range, as confirmed by Humphrey Instruments. While it is not clear exactly how these differences might affect the OCT images acquired, there is a substantial difference in quality between the OCT images reported and those which can be easily obtained with the commercially available Humphrey 2000 OCT. As for the other repeatabilities listed below, none has been reported in the literature.
2.2.2 Intersession and Long Term Repeatability

Intersession repeatability encompasses the variability that occurs when a patient is imaged over different sessions. That is, when a patient is allowed to get up and be repositioned between imaging sessions. Sources of this variability include all the causes of within-session repeatability and some new ones as well. For example, the patient’s positioning and machine settings could differ between the two different sessions. Thus, even for small time separations between measurement sessions, intersession variability is an important concern.

When the time separation is large, the patient’s retina itself may change between visits. Before OCT measurements can aid patient care decisions, the normal, long term variation of retinal thickness must be determined to the precision of OCT. Previous variability studies [2, 9, 10, 11], have all measured variability over scans that were either obtained within a single scanning session, or, in the last case, between sessions separated by just a few hours. These studies are presumed to capture the variability inherent in the OCT system itself, since retinal thickness is presumed to be stable over such short time periods. However, the stability of retinal thickness measurements over a day, a month, or longer periods has not been documented to the 10 μm precision of OCT. Normal retinal thickness is presumed to be stable over these periods based on appearance with slit lamp biomicroscopy; however, OCT is capable of far more precise measurements. Many other ocular physiological measurements such as intraocular pressure, corneal hydration, and even aqueous composition are known to have small diurnal and monthly variations. Even small retinal variations, if they exist, could greatly affect the precision of thickness measurements over time.
periods that are of clinical interest. Patient follow-ups, after all, normally occur over periods of days, months, or years rather than within a few hours.

2.2.3 Repeatability with Power Range and Dilation State

It is currently unclear how to interpret the reported measurement variability given the fact that OCT scans from the same subject can appear very different when acquired under different scanning protocols. Figures 2.1A and 2.1B show two such scans acquired during a single sitting, and the displayed images captured from the screen of the Humphrey 2000 OCT are similar to what the OCT operator would view as they were acquired. Such apparent differences are disconcerting as it is currently not known whether they result in different retinal thickness measurements; the retina in Figure 2.1A, for example, appears much thicker than in Figure 2.1B.

We also see effects caused by small pupils because the pupils shadow the OCT scanning beam and prevent it from returning a strong signal. These effects can be seen in Figure 2.2.

On a fundamental level, same-subject differences such as in Figure 2.1 occur because of differences in the interference signal intensity measured by the OCT at each point in the retina; such differences in signal intensity can be caused by variations in the scanning protocol. However, it is currently not clear whether study results using these different parameter ranges may be directly compared. No OCT image acquisition standards have been reported for pupil dilation or SLD power, nor is it currently known whether differences in B-scan signal intensity and qualitative appearance cause differences in measured retinal thickness. That is, it is not known if B-scans with different qualitative appearances contain different thickness information.
Figure 2.1: A-D. These images show how 2 B-scans acquired from the same subject but under different scanning protocols can appear very different, regardless of the display method. Parts A and B show the two scans as they would be displayed by the Humphrey 2000 during acquisition, while parts C and D show the same two B-scans using normalized, grayscale color maps. The higher intensity B-scan (A and C) was acquired with dilated pupils and a 750 W SLD power setting, while the lower intensity one (B and D) was acquired through undilated pupils and a 500 W SLD power setting. As described in a later chapter, the average scan intensities for the high and low intensity B-scans are 1075 and 422 intensity units, respectively, and their measured retinal thickness values are 275.0 μm and 281.3 μm respectively.

Figure 2.2: Here is an example of pupillary shadowing of the OCT image
2.3 Automatic Thickness Measurement

To characterize retinal thickness measurement variability, we need a practical means for measuring retinal thickness from OCT images. This method must be accurate and deterministic, so that a given image will always yield the same, correct thickness measurement. We wish to characterize the fundamental limitations of the Humphrey 2000 OCT system for measuring retinal thickness, and so all errors and variability in our experiments must be a result of the OCT images alone and not our technique for measuring them. Manual determination of boundaries for each image is not a satisfactory method, primarily because it is not deterministic and would introduce its own subjective variability. Moreover, it is not practical, particularly for the large series of images that must be obtained for sound variability studies.

The Humphrey 2000 OCT system comes with a software algorithm to display a plot of retinal thickness from individual scans. The algorithm works by detecting the retinal boundaries within the scan image and measuring the distance between them. However, this algorithm is also unsatisfactory. To begin with, it sometimes makes wildly incorrect estimations of boundary locations or fails to process scans at all. In these cases, the OCT operator has no recourse because no utility is provided for correcting algorithm errors. When the algorithm does work, it presents the thickness values on the computer screen in the form a graph whose data cannot be easily exported for statistical analysis. Without the ability to extract precise numerical measures from the plot, statistical operations such as regional thickness averages are difficult to perform. Samples of erroneous boundary detections by the Humphrey system are in Figure 2.3.
Figure 2.3: Here we have some samples of erroneous boundary detections by the Humphrey system. In the left image, note the boundary being displayed up in the vitreous humor, while on the right, the detected boundary lies well within the retina.

2.4 Video Tracking of Patient Eye Motion

Patient motion frequently occurs during OCT scan acquisition. These motions result from small postural adjustments and changes in fixation due to voluntary and involuntary saccades. The OCT operates at a very high magnification, and so even small eye motions can cause very large image disruptions. Large motions change the scan position to completely different retinal locations. In addition, the pupil margins may move into line with the scan beam, causing occlusions and shadows on the edges of the scan. The OCT operator will routinely reject scans which exhibit large motion or shadowing artifacts, and so such artifacts are not considered a problem. However, motion can occur without obvious artifacts in the OCT scans. For example, eye motions may cause only small breaks or jumps in the scan image. Alternatively, the eye’s position may slowly change over a series of scans, either by a slow gradual drift or by jumps that occur between scans that are retained. Thus by the time the operator saves the OCT scan, the scan location may be very different from the intended location.
The Humphrey 2000 OCT system does allow ocular motion to be directly viewed by means of a live video image of the ocular fundus as it is scanned. However, the system places the video image on a separate viewing monitor from the scan image. The OCT scan image requires constant attention during scan acquisition, and so even a skilled operator can only make occasional glances at the video monitor. The OCT also captures a still video image upon scan completion to record the final eye position. However, the image is captured a full second after the scanning is stopped, and so the same eye motions can occur during this interval. Furthermore, the halting of the scan beam can itself prompt subjects to change their gaze, causing the final picture to be inaccurate. Because of these problems, even gross, persistent changes in fixation are not always detected, and so one cannot be certain that a scan was obtained in the correct location.

The uncertainty of ocular position is problematic because it adds variability in thickness measurements and uncertainty in interpreting changes between scans. This is particularly true for measurements of localized abnormalities that do not feature clear landmarks such as the foveal pit. Another type of scan for which motion creates problems is a circle scan centered on the optic nerve head. These scans are intended to measure retinal nerve fiber layer (RNFL) thickness, which decreases rapidly with distance from the nerve head. In their study, Varma et al. studied cadaver eyes to measure normal RNFL thickness; from their plots the thickness decreases from approximately 350 $\mu m$ at the margin of the optic nerve head to 100 $\mu m$ at a distance of 750 $\mu m$ from the nerve head [12]. Further out, the RNFL decreases even more. The normal diameter for the nerve head is $1.68 \pm 0.13$ $mm$ vertically and $1.45 \pm 0.13$ $mm$ horizontally, and so a distance of 750 $\mu m$ represents a circle approximately twice
the normal nerve head diameter. This is the size circle used in many OCT studies of RNFL measurement.

Normal subject motion during OCT imaging can de-center circle scans, which may then radically change the RNFL thickness profile around the nerve head. Furthermore, nerve head scans require the subject to fixate far to the side. Such eccentric fixation requires constant attention, and so even normal subjects have increased motion. Even worse, subjects with advanced glaucoma have reduced visual fields and have even more trouble fixating. However, these are the subjects for whom accurate nerve fiber layer measurements are the most critical.
CHAPTER 3

BACKGROUND AND LITERATURE REVIEW

3.1 OCT Hardware

3.1.1 Fundamental OCT Principles

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 [13]. Figure 3.1 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 Figure 3.1 replaced by a mirror located exactly one meter from the beam splitter. The reference mirror in this thought experiment is initially placed a half meter from the beam splitter and then slowly moved outwards to a distance of one and one half meters.

The average signal power at the detector, $S$, can be modelled as the real portion of the cross correlation function between the reflected light from the imaging path and the reference path [14]. 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,
Figure 3.1: A schematic drawing of the OCT is shown here, emphasizing how it is essentially a Michelson interferometer. The outgoing light paths are solid lines, while reflected light is drawn as dashed lines.

\[ S(t) = \| \vec{E_i} + \vec{E_r} \|^2 \]  

(3.1)

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

\[ S(t) = E_i^2 + E_r^2 + 2E_iE_r \]  

(3.2)

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

\[ S(t) = (E_i + E_r)(E_i + E_r)^* \]  

(3.3)
where $E_i$ and $E_r$ are the complex phasor representations of the two light beams and the asterisk (*) represents complex conjugation. Eq. (3.3) can be expanded to

$$
S(t) = |E_i|^2 + |E_r|^2 + E_i E_r^* + E_r E_i^*
$$

$$
= |E_i|^2 + |E_r|^2 + 2 \text{Re} \left( E_i E_r^* \right)
$$

and the time average taken to yield

$$
\langle S(t) \rangle = \langle |E_i|^2 \rangle + \langle |E_r|^2 \rangle + 2 \text{Re} \left( \langle E_i E_r^* \rangle \right)
$$

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

$$
S(d) = c + 2 \text{Re} \left( \langle E_i(0) E_r^*(d) \rangle \right)
$$

where $c$ represents a constant offset. The rightmost term in Eq. (3.6) is the definition of a cross correlation [15]. 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 [16], as illustrated in Figure 3.2. 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, 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 Figures 3.3 through 3.4, we present sample OCT scans, and in Figure 3.3 we have labeled the inner and outer retinal boundaries along with other retinal landmarks. Also in Figure 3.3, the directions along which A-scans are obtained and then stacked are labeled. Figure 3.5 presents a sample A-scan. Figures 3.4 through 3.5 are all from the same
Figure 3.4: Yet another OCT scan is presented here to further demonstrate the possible variations. Figures 3.4A and 3.4B are different scans from the same person, while Figure 3.3 is from a different person. Note that in Figures 3.4A and 3.4B, 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.

Figure 3.5: Here we have a sample A-scan from Figure 3.4A, without any filtering. The DC offset is evident, as is the noise. The inner and outer retinal boundaries have been labeled.
subject, while Figure 3.3 is another subject. Though the general retinal shape is constant between scans, the details and shapes may differ wildly, even within the same subject. These differences are due mainly to noise and eye motion during scan acquisition, and do not represent actual retinal topology. 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 boundaries, particularly the outer retinal boundary near the fovea. Moreover, 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.

3.1.2 The Humphrey 2000 OCT System

The Humphrey 2000 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 dissertation are from a Humphrey system. The Humphrey SLD has center wavelength of 850 nm [16] 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 per A-scan and 100 A-scans per B-scan. 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
over-sampled, as the SLD half power width is 10 μm in the eye. For each A-scan, the reference mirror is moved through a distance of 2 mm.

During scan acquisition, the OCT operator can view the patient’s fundus on a separate video monitor. The fundus camera is black and white with a Volk lens as its objective to allow it to image the fundus. The camera’s field of view is approximately 30° x 20°. The camera is sensitive to infrared, so that it can see the reflection of the scanning beam on the surface of the retina. The camera normally images the retina with red illumination, but it is possible to switch it to a green illumination; this green illumination is referred to as red free. However, the green illumination can only be used to capture a single video frame, as it is too bright for subjects to tolerate for significant periods of time.

OCT scanning parameters that may be adjusted include pupil dilation state, pupil alignment with the OCT system, and SLD power. OCT studies have been performed with both dilated [9, 10] and undilated [2, 11] pupils and with SLD power levels ranging from 200 to 750 μW [9, 10, 17]. These parameters affect the light intensity reaching the retina and, consequently, both the interference signal intensity returned to the OCT and the appearance of the OCT scan. SLD power directly affects the light intensity incident on the retina, while pupil dilation causes indirect effects. Subjects with fully dilated pupils can be aligned with the OCT system more easily than those with small pupils; the alignment is important because even small misalignments can degrade OCT scan quality by causing shadowing artifacts, as in Figure 3.6, or scans that are darker overall.

OCT B-scans are displayed as pseudocolored images, whose colors denote the interference signal intensity from each location. The 172 colors range, in order of
Figure 3.6: This image displays artifacts (indicated by arrows) that can result from shadowing of the retina by the pupil margins. The result is that the retina is obscured in the shadowed region. This image is displayed by the Humphrey 2000 using a normalized color map.

decreasing signal intensity, through white, red, orange, green, blue, and black. During their acquisition, B-scans are displayed with an absolute color map, so that each color always represents the same, specific signal intensity range. For printing purposes, a normalized color map may be selected, whereby white represents the highest signal intensity within the scan and black the lowest, regardless of the actual signal intensity values. Figure 3.7A and B show a high and low intensity scan pair using an absolute color map, while 3.7C and D show the same scan pair using a normalized color map for comparison.

The raw signal intensity data can also be exported from the Humphrey system and displayed using a grayscale map [18]. Here the interference signal intensity for each pixel is displayed as one of 256 shades of gray, ranging from white for the highest intensities to black for the lowest. As with the Humphrey color maps, the grayscale map can be normalized. The use of grayscale maps avoids quantization artifacts and the artificial perception of boundaries that may result from the use of color.
In particular, the Humphrey color map contains 172 colors, but Humphrey images are generally perceived as having only a few colors with distinct boundaries between them. This is the same effect seen with rainbows, which are generally perceived to have only the traditional red, orange, yellow, green, blue, indigo, and violet, despite the fact that they contain hundreds of colors in a smooth transition from red to violet.

The Humphrey system comes from the manufacturer with software tools for measuring retinal thickness. However, there is no literature describing precisely how their system works. However, we were fortunate enough to receive a technical bulletin from the manufacturer. Their system studies each A-scan independently and models the retinal boundaries as peaks. The individual A-scans are heavily smoothed and the algorithm studies each A-scan in an outwards direction starting from the innermost end. The inner retinal boundary, then, is modelled as the first, large peak in the A-scan. The algorithm then detects the middle of the photoreceptor layer as the deepest trough encountered, and the outer retinal boundary is detected as the next large peak after the photoreceptor layer trough. There are provisions for detecting breaks and filling them in with information from the surrounding A-scans, but this is not well documented.

The Humphrey system often gives reasonable boundary determinations for the images; examples of these are in Figs. (3.11-3.8), where we have simply shown the images of the OCT scans with the output boundaries displayed. However, a significant problem is that the Humphrey system frequently makes errors in its boundary determinations. A severe case is in Fig. (3.9). In those images, the Humphrey system has placed the boundaries well off the retina and within the vitreous humor. For a few images, no detected boundary is displayed; we have not shown these images as there
Figure 3.7: This figure shows another example of how two B-scans, acquired from the same subject, can appear very different regardless of the color map used for display. The differences persist even with the normalized color maps, indicating that they are not simply a result of scaling. The high intensity B-scan is on the left, and the low intensity B-scan is on the right. Parts A and B use the Humphrey absolute color map, C and D use the Humphrey normalized color map, and E and F use a normalized grayscale color map. As described later in the paper, the average scan intensities for the high and low intensity B-scans are 1023 and 652 intensity units, respectively, and their measured thickness values are 283.3 μm and 277.6 μm respectively.
Figure 3.8: This figure shows more examples of acceptable performance of the Humphrey boundary detection system.

is nothing to show other than an OCT scan with no boundaries superimposed. The errors are not always so dramatic, of course, but they frequently occur. Examples of these more typical errors are in Fig. (3.10). These are errors that are large enough to be objectionable and disrupt measurements. Because the Humphrey data are locked within the computer system, there is no easy means to redirect the boundary detector or otherwise correct errors when they occur. It was our frustration with these shortfalls that was the impetus for our work in designing a better OCT boundary detection system.

The detected boundaries are then used to measure retinal thickness, and this thickness is presented to the user as an onscreen plot. Neither the image of the thickness plot nor the detected boundaries can be directly exported; until recently,
Figure 3.9: This figure shows examples where the Humphrey boundary detection system completely failed.

Figure 3.10: This figure shows examples of errors that typically occur with the Humphrey boundary detection system. These errors are large enough to be objectionable and disrupt measurements.
the only option was for the user to perform a screen capture. One such screen capture is presented in Fig. (3.11). The upper left quadrant displays the OCT image with the retinal boundaries as white lines. The bottom left shows the plot of retinal thickness. The upper right shows a captured frame from the video camera that images the subject's retina as it is scanned by the SLD, and the lower right has patient information and scan parameters. One shortfall of the Humphrey system is immediately apparent. The plot has no grid lines to help make measurements from it, nor are the data used to generate the plot directly available for export. Recently, this problem has been addressed by the distribution of a utility that allows the user to export the plot data in the form of a text file. Though awkward, this utility is certainly an improvement because, previously, precise numbers could not be determined for retinal thickness.

The Humphrey OCT also has a facility for measuring RNFL thickness. The Humphrey algorithm is subjective in that it marks the outer RNFL boundary for a given retinal location as the location where the OCT signal is attenuated by a certain number of dB. This number is empirical and is based on internal Humphrey studies [19], but similar techniques are reported in the literature [7].

3.2 Prior Clinical Studies

OCT has been used to image a multitude of retinal pathologies. Macular lesions associated with optic nerve head pits [20], epiretinal membranes [21] central serous chorioretinopathy, ARMD [22], choroidal neovascularization [22], and diabetic macular edema [10] are just some of the pathologies that have been studied using OCT.
Figure 3.11: Here is a sample screen capture for the Humphrey system, showing the information that is given to the user.
For many of these studies, the OCT's fine in vivo resolution has given insight into disease mechanisms, particularly macular holes [23], macular edema [10, 24, 25, 26, 27], ARMD [22], and various choroidal pathologies [28, 29, 30, 31]. In these cases, the OCT is being used to obtain qualitative information about the anatomical structure of the retina.

OCT can be used to obtain quantitative information, as well. Several articles report the measurement of foveal and parafoveal macular thickness with OCT [27, 32, 33], and the short term measurement variability has been assessed to be on the order of 10 μm [2, 9, 10, 34]. These variability studies, however, all measured within-session repeatability and all used prototype machines rather than the Humphrey 2000 OCT. Quantitative information may be important in diseases such as diabetic macular edema, where the quantitative thickness of the retina is an indicator of the level of disease and the efficacy of treatment. Another pathology for which measurements are important is glaucoma. Here the thickness of a portion of the retina is measured, namely, the retinal nerve fiber layer. These measurements are traditionally taken in a circle around the optic nerve head. Such a view allows the clinician to detect defects that might affect any portion of the visual field. This dissertation is dedicated to investigations which are particularly suited for improving the clinical measurement of retinal thickness and RNFL, and so these two diseases are likely beneficiaries of this research.
3.3 General Ophthalmology

3.3.1 Diabetic Macular Edema

Macular edema is the most common cause of vision loss in diabetics [4]. The condition is widespread; one can find edema within one disc diameter of the fovea in 9% of diabetics [35]. This prevalence is particularly significant because untreated patients have a greater than 50% chance of losing one or more lines of vision within two years [36]. Clinically, macular edema is defined as a thickening of the macula as seen by slit lamp biomicroscopy or stereo fundus photography [4]. In advanced cases of edema, one can see cyst-like spaces formed within the retina [4].

DME exists in three forms: focal, diffuse, and combinations of these two types [35]. Focal edema is associated with focal areas of fluid leakage as seen by fluorescein angiography [37]. These leakage points correspond to vascular lesions such as microaneurysms or dilated capillaries [37]. Often one will see circular rings of hard exudates, also called lipid, surrounding these leakage points [37]. The hard exudate is formed when plasma containing lipoproteins leaks from the vascular lesions; as the fluid reaches healthy tissue, it is resorbed, leaving the lipoproteins behind to collect in the outer retinal layers [4].

For treatment purposes, macular edema is classified into clinically significant macular edema (CSME) and non-CSME. CSME is defined by one of three conditions: (1) retinal thickening and/or hard exudes adjacent to thickened retinal regions involving the center of the macula; (2) retinal thickening and/or hard exudate adjacent to thickened retinal regions extending to within 500 \( \mu m \) of the center of the macula but not involving the center; or (3) retinal thickening involving one disc area or more of
retina, part of which is within one disc diameter of the center of the macula, but with neither thickening nor adjacent hard exudates within 500 μm of the center [5].

For macular edema that could be classified as CSME, a comprehensive study called The Early Treatment Diabetic Retinopathy Study (ETDRS) determined that treatment with a laser was beneficial, whereas it was not beneficial for non-CSME [5]. Two forms of treatment, focal and grid laser surgery, were found to reduce the incidence of moderate visual loss (doubling of the visual angle) by 50 to 70% over three years [5]. Both forms of treatment cause irreversible damage to retinal tissue; thus the treatment is only applied in cases where it is felt the potential benefit will outweigh the loss due to damage.

The ability of OCT to give numerical measurements of retinal thickness may be of use in treating CSME. To begin with, the laser surgery does not always reduce the edema and repeat treatments are sometimes called for. The ETDRS derived protocol requires re-treatment of persistent edema threatening the center of the macula, assuming it can be done safely [5]. Currently, ophthalmologists must rely on their chart notes, photographs, and memory to determine whether the retina has changed between visits. However, it is sometimes difficult to judge whether edema has increased, decreased, or not changed at all. One ophthalmologist estimated in conversation that he could probably detect changes of 25%. Another potential benefit was suggested by a recent analysis of the ETDRS data. This study reported that the initial degree of central macular thickening and initial area of macular thickening are the strongest predictors of surgical outcome [38]. These gradations were performed subjectively; it is not known if precise, numerical measurements could yield new, objective treatment protocols. Finally, accurate and precise measurements may detect slight retinal
changes, indicating the progression or incidence of disease before it can be detected by current methods.

### 3.3.2 Glaucoma

Glaucoma is the leading cause of preventable blindness in the United States [6]. Two million Americans have glaucoma, with 80,000 of these having been blinded from the disease. The mechanism of glaucoma is impaired outflow of aqueous humor from the anterior segment of the eye, resulting in elevated pressures and cell death [6]. The treatment of glaucoma, then, involves preventing damage by controlling the ocular pressure, either by decreasing aqueous production, or by improving its drainage. Primary open angle glaucoma is the most common form, and is often undetected until severe visual field loss is already present. This form of glaucoma causes slow, progressive destruction of the retinal nerve fiber layer (RNFL) and inner nuclear layer, resulting in a progressive loss of vision [6]. Once a deficit in the visual field is present, the damage is irreversible; thus, early detection is crucial [7]. One of the earliest detectable signs is destruction of the RNFL. This destruction can be localized or diffuse, and most glaucoma patients feature both [7].

OCT offers the potential to measure RNFL thickness directly; OCT measures of RNFL thickness have been shown to correspond to the presence or absence of glaucoma. Mistleberger et al. studied 17 normal eyes, 23 eyes with elevated ocular pressure but no glaucoma, and 38 eyes with glaucoma [3]. They found that the glaucomatous eyes had significantly thinner RNFL thickness measurements. Pieroth et al. used OCT to detect focal defects [7]. Their study group was 25 eyes of 19 patients. For each scan they derived a plot of RNFL thickness at every A-scan. This
plot was then normalized by the average thickness value for that scan and the variation of thickness values in that scan. A normative data base was constructed, and focal defects were defined as large, negative deviations from this normative database for at least three adjacent A-scans.

3.4 Relevant Computer Vision Background

3.4.1 Boundary Detection in OCT images and Noisy Images in General

OCT B-scans

A preliminary version of the retinal boundary detection work in this dissertation appeared at The 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [39]. To the best of our knowledge, only one other computer vision work has been published on OCT images to date. This work, by George et al [40], 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.

Noisy images in general

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 [41], such as [42, 43]. The authors in [42] 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. In [43] they use 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, among others, Sha'ashua and Ulman [44].

A variation of snakes that has been used to find feature boundaries in medical images is a level set approach [45, 46]. 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 [47] but by improved low level edge detectors [48, 49] as well. In another technique, the edge detector scale is adjusted based on local image variability [50]. Statistical approaches have been proposed, such as Bayesian decision making based on local variations in mean intensity [51] and detection of edges using finite mixture distribution analysis [52]. In [52] the authors 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; see, for instance [53, 54, 55, 56]. 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, [56] 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.

3.4.2 Feature Detection Tracking in Fundus Images

Detection of blood vessels in fundus photographs

Chaudhuri et al. explore one possible means of detecting retinal blood vessels in their work [57]. They used matched filters, also known as templates. They noted that the intensity profile of blood vessels in their images had a Gaussian shape, and designed templates to match this. Their template for detection of vertical vessels was 15 × 16 pixels. The top and bottom three rows were zeros, and the other rows were identical Gaussian profiles, offset so that the sum of each row was zero. Each Gaussian profile extended for three standard deviations to either side of the mean. They determined that using a single standard deviation value for all templates, corresponding to a medium caliber vessel, was sufficient. Using templates with a range of standard deviation values to capture large and small vessels did not give noticeable improvements.

Similar templates were generated for other vessel orientations by use of a 2 × 2 rotation matrix, where each rotated template remained 15 × 16 pixels in size. They concluded that 12 equally spaced templates were sufficient to obtain good results, and that additional template orientations did not yield significant improvement. The fundus images were convolved with each of the 12 templates, and each pixel in the output image was given the maximum of its responses to the different templates. Their images were 480 × 512, and appear to comprise twice the viewing angle of
the OCT camera. Also, their images are captured with red-free light, and so have much higher contrast than the OCT video images. Their results appear very good. However, their output images locate the entire blood vessel, and so some sort of thinning algorithm is necessary to reduce the vessels to single contours.

Can et al. also developed a system for detecting blood vessels in retinal images.[58] Their system was designed to detect blood vessel bifurcations in retinal images taken from a high resolution camera (1024 × 1024) using high contrast, red-free lighting. Their approach was to iteratively trace out the vasculature using edge detectors to detect the local vessel contours. They initially locate a set of "seed" points that are likely to fall on a blood vessel. The points are chosen based on local image characteristics and an initial, crude search. A set of edge detectors are then applied to each point. The detectors are optimized to detect parallel vessel edges, separated by some distance \( d \). Different separations are tried along a set of nine radial directions, and the maximum results are noted and compared with the results of neighboring points. A determination is then made whether a point lies on a vessel, and, if so, in which direction the vessel is oriented. The next further point along the vessel is then analyzed in the same way, and a linked list is created for every vessel grown in this iterative fashion. The initial selection of seed points captures many points along most blood vessel segments, and so the algorithm determines when two points represent the same region of a blood vessel, and when linked lists intersect to represent either a single, longer blood vessel or a bifurcation. This avoids the thinning process necessitated by Chaudhuri's algorithm.

Detection of the bifurcations are in fact the goal of the algorithm presented by Can et al., and these bifurcations are used in another paper they write for finding
correspondences between retinal images. They choose these features over a template based approach like Ott's[59] because their images may have only small overlap. In particular, Can et al. describe how modelling the eye as a quadratic surface allows good matches between corresponding features in different views of the retina [60]. In this work, they are trying to form a mosaic of retinal images to create a wide field view of the entire retina. They then need to be able to match smaller field views to the correct location in the mosaic. They model ocular motion as a rotation about two axes, both parallel to the image plane, and note that the rate of rotation can be as high as 180 degrees per second. No significant rotation is assumed to occur along the optical axis between the image plane and the eye.

As Can et al. explain, a hierarchy of models can be used to describe the image distortions caused by eye movement. One can go from a simple, zero-order translation model, to a first order affine transform, and then to a second order model which takes the quadratic retinal shape into account. Their approach to finding the transformation between two images is to first assume a zero order model and find the translation vector that minimizes the error between feature correspondences. They then use this result to estimate an affine transform, and then use that to estimate a quadratic transform. An affine transformation model usually results in a 4-5 pixel error between feature locations in corresponding images, for 1024 by 1024 pixel images. The quadratic transformation, however, reduces this error to 1.1 pixels, on average. Robust statistical methods are used to determine the affine transform and the best quadratic transform. However, the field of view for their camera is approximately twice that of the camera in the OCT. The sample mosaic images they show feature
shifts that are 75 to 100 percent of the field of view of the OCT. Thus with OCT images, we discovered that the zero order model was sufficient.

The work of Can and his colleagues is geared towards the development of a system to automatically track eye motions during laser surgery [61], and appears superficially to be similar to the video tracking problem associated with OCT. However, it should be noted that the images they show in their paper and on their web page are obtained with red free illumination, and so again, they can locate far more features per image. Also, as will be explained later, there are concerns associated with tracking the laser spot in the OCT. Thus, while their work can serve as a guide for addressing certain issues, substantial modifications were necessary to develop a similar tracker for the OCT.

Tracking of eye motion

A relatively small portion of the computer vision literature appears to be dedicated to the problem of ophthalmic images. Ott and Daunicht developed a method to calculate ocular movements based on changes in a video image of the fundus (back of the eye) [59]. Their goal was to understand these movements for physiological evaluation purposes. They intended their work to be used with images from a scanning laser ophthalmoscope (SLO), which gives highly magnified, high contrast views of the retina. They modelled the eye as a rigid, spherical body having a fixed rotation center and known radius, and assumed that the eye could rotate about arbitrary axes. They used a pinhole camera model.

From the shifts retinal features make in the image plane, Ott and Daunicht provided formulas for calculating the Euler angles for the corresponding ocular rotations. They suggested using a disk shaped foveal region in the initial image as a template.
An equally sized foveal region in subsequent images could then be compared to the reference template to calculate the ocular rotation. They also recognized that eye motions would distort the initial, disk shaped region into a tilted ellipse in subsequent images and described a means to compensate for this. Their methods depend on being able to match pixel intensities in the reference template with the template in the new image. This approach may work well for SLO images, which have high contrast. However, OCT video images suffer from low contrast. Also, pixel intensities will not remain constant between images because the scanning beam changes the illumination.

3.4.3 Eigenimage Techniques for Object Recognition

Basic overview and theory:

Sirovich and Kirby [62, 63] described the use of principle components analysis to describe faces. Turk and Pentland expanded the concept to the detection and recognition of faces in images [64]. We will explain the fundamental principles of eigenimage analysis, as in [64, 65]. First of all, one must conceptualize images as vectors. Hence an $N \times N$ pixel image $I$ can be thought of as a vector in an $N^2$ dimensional space. Say one considers an object class, $\mathcal{C}$, and a set of $M$ sample training images, $\{\tau_C^i\}$, such that $\tau_C^i \in \mathcal{C} \forall i$ (the superscript $i$ indexes the $M$ training $\tau_C^i$). We will be referring to the $\tau_C$ as vectors or images, where appropriate. For many real-world classes, such as “fire engines”, “faces”, or even “optic nerve heads”, the $\tau_C$ will not be randomly scattered in $\mathcal{R}^{N^2}$ because all the $\tau_C$ will have some common, underlying structure. They can thus be imagined to occupy some subregion of $\mathcal{R}^{N^2}$.
This subregion can be approximated by a subspace, \( \Phi_C \) that is a mathematically optimal approximation. The method we use to find \( \Phi_C \) is principal components analysis (PCA), also known as the Karhunen-Loeve expansion. This method finds vectors that capture most of the energy in \( \{ \tau_C \} \) by finding the eigenvectors of the covariance matrix of \( \{ \tau_C \} \). By "capture most of the energy," we mean that we minimize the average distance between each \( \tau_C \) and its closest point in \( \Phi_C \). This is important, because we ultimately represent each \( \tau_C \) by its projection into \( \Phi_C \). To implement PCA, we first find the covariance matrix \( K \) as

\[
K_C = \frac{1}{M} \sum_{i=1}^{M} \tau_C^i (\tau_C^i)^T
\]  

Then the eigenvectors of \( K_C \), \( \{ \phi_C^1, \ldots, \phi_C^M \} \), span the eigenspace \( \Phi_C \) and are thus a basis. If \( M \geq N^2 \), then there are at most \( N^2 \) eigenvectors with non-zero eigenvalues; otherwise, there are at most \( M \). The calculation of the eigenvectors is made easier if one takes advantage of singular value decomposition (SVD); in cases where \( M \ll N^2 \), we can find the eigenvectors of an \( M \times M \) matrix rather than an \( N^2 \times N^2 \) matrix. To do this, first define \( X = [\tau^1 \cdots \tau^M] \) and rewrite Eq. 3.7 as

\[
K_C = XX^T
\]  

In that case, if \( r \) is the rank of \( X \), then \([65]\)

\[
X = \sum_{k=1}^{r} \sqrt{\lambda_k} u_k v_k^T
\]  

The \( u_k \) and \( v_k \) are, respectively, left and right singular vectors of \( X \), and the \( \lambda_k \) are the non-zero eigenvalues of \( K \). The \( u_k \) are the eigenvectors of \( X^TX \), which is
$M \times M$; this observation is useful because the $u_k$ are the eigenvectors of $XX^T$, which is $N^2 \times N^2$, and

$$u_k = \frac{1}{\sqrt{\lambda_k}} X v_k$$

(3.10)

Therefore by finding the $v_k$ we can greatly reduce the necessary computations.

The magnitude of the eigenvalue, $\lambda_k$, corresponding to $\phi_k^C$, indicates the relative importance of $\phi_k^C$, because it represents the mean squared energy value of the set $\{\tau_k^C\}$ that is captured along its direction. If $\tau_k^C$ represents the projection of some $\tau_C \in C$ onto $\phi_k^C$, then the sum

$$\sum_{\tau_C \in C} ||\phi_k^C \tau_C - \tau_C||$$

(3.11)

is minimized for $\phi_k^C$ corresponding to larger $\lambda_k$. As further explained [65], $\lambda_k$ indicates the variance of $\{\tau_k^C\}$ along the direction of $\phi_k^C$. If we wish to choose a most representative subset of $m$ elements from $\{\phi_k^C\}$ to represent the $\{\tau_k^C\}$, then choosing the $\phi_k^C$ according to their eigenvalues is mathematically most efficient. So, if we call the eigenspace constructed from these $m$ vectors $\hat{\Phi}_C$ (where the “hat” operator indicates an approximation), then $\hat{\Phi}_C$ is the $m$ dimensional space which minimizes the average error between $\{\tau_k^C\}$ and the projection of $\{\tau_k^C\}$ onto $\hat{\Phi}_C$. Note then that eigenvectors with zero eigenvalues are not desirable because they capture none of the variance or energy of the $\tau_k^C$. By constructing $\hat{\Phi}_C$, we thus have a subspace that elements of $C$ project onto very well, and, hopefully, $\tau \not\in C$ do not.

This is how we then use the eigenspace, $\hat{\Phi}_C$ to detect objects of class $C$ in images. In particular, say we have an unknown image, $\tau$ we wish to classify as being in $C$ or not. We project $\tau$ into $\hat{\Phi}_C$ to generate the projection $\hat{\tau}$, and measure the error,
\( \epsilon = \| \hat{\tau} - \tau \| \). If \( \epsilon \) is below some threshold, we declare \( \tau \in C \), otherwise, we declare \( \tau \not\in C \). In the case where we have a larger image \( I \) that is much larger than the training images, \( \tau^*_C \), we can investigate all possible subregions \( \tau^* \) contained within \( I \). Each \( \tau^* \) is projected into \( \Phi_C \) and \( \epsilon \) is calculated. If \( \epsilon \) is below the threshold for any \( \tau^* \), then that \( \tau^* \) is assumed to contain an object of class \( C \).

The eigenimage technique can be expanded to distinguish not only between images of class \( C \) and non-class \( C \), but also between different subclasses of \( C \). In particular, say that \( C \) is composed of several subclasses, i.e. \( C = \bigcup_j C_j \), all of which are represented in \( \{ \tau^*_C \} \). In particular, say that \( \tau^*_C \) is a training example of subclass \( C_j \). Then each \( C_j \) will project into a different locus of points in \( \Phi_C \), where the center of the locus for \( C_j \) is \( p_j \). Thus for an unknown \( \tau \), we first project it into \( \Phi_C \) to generate \( \hat{\tau} \); from the projection error, \( \epsilon \), we determine if \( \tau \in C \). If so, we then determine which \( p_j \) is nearest to \( \hat{\tau} \) and conclude that \( \tau \) is an element of the corresponding \( C_j \).

To help with the between-subclass differentiation, Turk and Pentland [64] advocated first finding the average \( \tau_C^i, \tau_C \)

\[
\tau_C = \frac{1}{M} \sum_{i=1}^{M} \tau_C^i
\]  

and then removing \( \tau_C \) from all training images, so that for each \( \tau_C^i \) we generate \( \tilde{\tau}_C^i = \tau_C^i - \tau_C \). In that case, we could calculate

\[
\tilde{K}_C = \frac{1}{M} \sum_{i=1}^{M} \tilde{\tau}_C^i (\tilde{\tau}_C^i)^T
\]  

The benefit of removing \( \tau_C \) is that it removes the common term from among the \( \tau_C^i \), so that we only model the differences and thus create a wider separation (relative to their magnitude) of the \( \tau_C^i \).
Many reports of eigenimage based systems exist in the literature. For example, Watta et al. [66] developed a system for estimating the position of an automobile driver. They had sample images of test subjects in one of five different poses, and constructed an eigenspace for all their images. They projected an image in question into the eigenspace and found which training samples it projected closest to. It was then assumed that the image had the same pose associated with the nearest training samples.

Problems with eigenimage based techniques and solutions

Eigenimage techniques tend to have problems coping with changes in object scale within images, as well as differences in lighting and pose. Turk and Pentland[64], for example, found detrimental performance changes from variations in lighting, position, and scale. The most sensitive parameter was scale; varying the image size by a factor of 16:1 reduced performance by half. Much work has appeared in the literature since then, and has addressed means to improve the between-class differentiation of the eigenimage techniques as well as the problems of lighting, scale, occlusion, and pose.

Pentland et al. [67] addressed the problem of pose sensitivity by building a very large database of 7562 images from 3000 individuals. These images captured the individuals under different lighting, facial expressions, and poses. They then tried to address the problem of pose by comparing two different approaches. Normally, with \( n_s \) subjects at \( n_o \) orientations, one can build one eigenspace \( \Phi \) encompassing all \( n_s n_o \) possible subject-orientation combinations. A given \( r \) is then projected into \( \Phi \), and the distance is measured to each of the \( n_s n_o \) loci to determine subject identity and orientation. Alternatively, one can generate \( n_o \) separate eigenspaces, one for each orientation. One then determines which \( \Phi_o \) creates the least projection error.
to determine orientation; afterwards one looks within the \( n_s \) loci within the optimal \( \Phi_0 \) to determine subject identity. No large difference was observed between using either technique. In our case, we did not have real differences in orientation, since the curvature of the eye was not a significant influence on the nerve head appearance. We did have different lighting possibilities, but the number of possible variations precluded constructing separate databases for each possibility. Moreover, we really have only one subject, namely the nerve head and are not interested in identifying the illumination.

Li and Lu did work [68, 69] to address differences in pose and lighting. Their work addresses how one finds the best match for an object in \( \Phi_C \). The classic method is to look at the locus of points for each subclass of object to be recognized, \( C_j \), and find the center \( p_j \). Then an unknown image, \( \tau \) is projected into \( \Phi_C \), and one finds the closest \( p_j \). Instead of this nearest center approach, they suggest modelling the locus of points from each \( C_j \) as a line in \( \Phi_C \) instead. Then one finds the distance from \( \tau \) to the nearest line. The class \( C_j \) corresponding to the line is then assumed to be the identity of \( \tau \). The use of this line instead of the locus center allows variations in the image, due to lighting and position, to be modelled without having to have explicit examples in the training images. Thus use of the nearest line reduced the misclassifications in their evaluation by half.

Purnell et al. [70] address one interesting aspect of eigenimage approaches to facial recognition—that different races have different facial characteristics. They created different databases of Caucasian and African-American training images, and compared recognition performance on each database. They also noted that to compensate for changes in lighting, normalizing images to zero mean was helpful. They
found that they got best results for both groups when they used a combined eigenspace rather than eigenspaces trained to each specific group. They attributed this to the normalization of the images prior to evaluation. Hence, once normalized, the Caucasian and African-American images did not vary much among themselves. In our problem, the nerve head appearance certainly varied between subjects, but then we also normalized our images.

Bischof et al. [71] addressed the problem of occlusion and scale by generating a novel approach to eigenimage detection. They are also one of the few groups not to use facial images. Rather than projecting an entire image $\tau$ into an $m$-dimensional eigenspace $\Phi$, they use a random sampling technique to pick $m$ sample points from $\tau$. They estimate the coordinates of the projection, $\tilde{\tau}$ in the eigenspace using the $m$ sample points and the $m$ basis eigenvectors by solving the implied set of $m$ linear equations. They do this for many sets of $m$ samples in $\tau$, which allows them to find a consensus that is robust in the presence of outliers. Also, if $\tau$ shows one or more objects that are partially occluded, this technique allows the individual objects to be identified, since the solutions will cluster around one or more possibilities. Another benefit of this approach is that the coordinates in $\Phi$ are invariant with mutual scale changes of $\tau$ and the eigenimages $\phi_i$. Hence a coarse to fine scale search could be run on a larger image to efficiently find an object, since at each iteration the coordinates in $\Phi$ would be refined. This approach also allowed image scale in $\tau$ to be estimated quickly, by seeing how the projection error varied as the image scale was adjusted.

This approach was intriguing, but not necessary and possibly not applicable in our domain. This approach seems more suited to detecting or distinguishing from among a set of high contrast object classes. The test images shown were high contrast, so for
most points picked within an object, useful information could be deduced from how it projected onto the various $\phi^i$. With the nerve head images, they tend to be bland with only low contrast features on their interior. Also, there are no other objects to occlude the nerve head, so occlusion is not a concern. Though the SLD spot does, in effect, occlude the nerve head, our masking technique and the large number of sample images used compensated sufficiently. This method was of some interest because of its ability to handle different scales well; however, our method performed adequately and so we did not investigate any changes.

**Linear discriminant and other techniques**

Some work has been published regarding the use of the linear discriminant analysis, or Fisher's linear discriminant instead of PCA and the resulting eigenimages. The idea is introduced by Belhumeur *et al.* [72] and used by Meng and Nguyen [73] and Martinez and Kak *et al.* [74]. The technique was introduced as a means to improve the ability of systems to distinguish among the many different subclasses $C_j$ that comprise the class $C$. The basic idea is that PCA finds eigenvectors that maximize the variance within $C$. However, this can be due to maximizing the variance within the individual subclasses as well as the variance between subclasses. For the purpose of distinguishing individual subclasses, however, we would like to find vectors which maximize the variance between subclasses and minimize the variance within classes. One set of vectors which does just this are the linear discriminants, and Martinez and Kak explain how to calculate them. Say that $r_j^i$ denotes the $i^{th}$ training image of subclass $C_j$, that the mean $r_j$ is denoted $\bar{r}_j$, and that the mean of all $r$ is $\bar{r}$. In that case, we define the within class covariance matrix, $S_w$ as
\[ S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (\tau_i^j - \bar{\tau}_j)(\tau_i^j - \bar{\tau}_j)^T \]  

(3.14)

where \( c \) is the total number of classes, and \( N_j \) is the number of \( \tau_j \) in class \( j \). The between class covariance matrix, \( S_b \) is given by

\[ S_b = \sum_{j=1}^{c} (\bar{\tau}_j - \bar{\tau})(\bar{\tau}_j - \bar{\tau})^T \]  

(3.15)

The linear discriminant vectors are then given by the non-zero, generalized eigenvectors of \( S_w^{-1} S_b \). There are at most \( c - 1 \) non-zero, generalized eigenvectors, and we need at least \( N^2 + c \) training images to guarantee that \( S_w \) is invertible. (Recall that we assume \( \tau \) is an \( N \times N \) pixel image.) Since this is not always possible, one solution is to use PCA to represent the \( \tau \) in a smaller dimensional space, and then use a linear discriminant to separate them there.

Belhumeur et al. [72] showed that the linear discriminant could be used to recognize a large number of faces under diverse lighting conditions better than PCA derived eigenimages could. Meng and Nguyen used a linear discriminant to perform detection, and so their work is relevant to this problem. They had 480 training images of faces and cluttered background. They considered two classes: face and non-face, and created sample training images of each. From these sample images, they calculated the one dimensional linear discriminant, and found it to have a false positive rate for finding faces of 7.3% in their evaluation set of 286 images. We considered using a linear discriminant, but rejected it in favor of the dual eigenspace technique, described in the chapter on video tracking. Finally, Martinez [74] noted that linear discriminants gave superior performance than PCA based techniques, so long as one has a large number of samples of each data subclass. If one only has a
few samples of each subclass, however, then the linear discriminant can in fact give worse results.

Other recognition methods

Zhang et al. [65] did a comparison between eigenimage, elastic matching (deformable templates), and neural network techniques for face recognition. They tried all three methods on a combined set of four different image databases and compared performance. All databases contained mostly frontal views of faces, and were adjusted to maintain a similar scale. However, there were some positional and lighting differences among the facial images. The deformable templates did best overall, with a correct identification of 80% of the images in the combined databases. The eigenimage technique only achieved a 60% correct classification, mainly because the four individual databases had large variations in illumination among themselves. Hence the differences between individual images and the eigenspace $\Phi_C$ could be dominated by the differences in illumination rather than among the faces themselves. This exposes one of the fundamental weaknesses of eigenimage techniques, namely that they are unable to cope with large variations in illumination. The neural network technique did far worse than either of the other two. Neural network performance is bounded by eigenimage performance, and so one cannot expect a network to do better. Moreover, the network design is not transparent; instead, one trains a network and settles for whatever the final connection weights are. If performance is not optimal, it is not clear how to affect things by modifying the weights.
CHAPTER 4

VARIABILITY BETWEEN AND WITHIN SESSIONS

4.1 Materials and Methods

4.1.1 Equipment

As we said above, OCT is based on the principles of laser interferometry [13, 75]. The output from a low power infrared super-luminescent diode (SLD) is split so that one half travels down a reference path and the other half travels into the eye and reflects back [75]. The reference and reflected beams are recombined, and heterodyning allows the optical interference of the two beams to be detected [76]. The use of a low coherence length laser allows only the reflections from a narrow region of the retina to interfere with the reference beam [76], giving the theoretical resolution of less than 10 \( \mu m \) that has been reported [9]. This region is scanned through the retina at a single point to generate an A scan, and this scanning point is moved over the retina to generate a B scan [13]. The B-scan images generated by OCT represent infrared reflectivity within the tissue and are thus analogous to ultrasound B-scans, which depict ultrasound reflectivity [13].

The Humphrey 2000 OCT system used in this study is a commercial implementation of the prototype systems that have been previously reported. It uses an 850 nm SLD source with a patient exposure range of 200-750 \( \mu W \). Other studies have
reported machines using SLD wavelengths of 830-850 nm, with incident power ranges of 200-1000 μW.

4.1.2 Subjects

Twenty six volunteers (13 male and 13 female), ranging in age from 20 to 52 years (mean ± SD = 31 ± 9) participated in this study. The volunteers were selected from research associates and clinic staff. Exclusion criteria included subjects with previous known retinal pathology, as elicited from history. Informed consent was obtained, and the protocol was approved by The Ohio State University Biomedical Sciences Institutional Review Board.

4.1.3 Pupil Measurements

All subjects in this study were imaged through undilated pupils. Pupil size can vary considerably, however, between normal subjects. Thus, in order to assess the applicability of this study's results to other subject populations, each subject's pupil was measured both in dark-adapted and light-adapted states. A pupil gauge was held on the subject's forehead while the subject fixated on a distant object. Pupil measurements were made in a darkened room using a transilluminator held obliquely from below to visualize each pupil before it contracted. The transilluminator was maintained shining into the pupil for the constricted pupil measurements. This method was selected because it is used in routine practice. Estimates of pupil diameter were made to the nearest 0.5 mm. All subjects but one had their pupil measurements performed by the same investigator (SEK).
4.1.4 Refractive Error

Refractive error was assessed by either using an automatic refractometer or by assessing the prescription from the subject’s glasses, if they were obtained within the last 12 months. The automatic refractometer used was the Automatic Refractor Model 585 (Allergan Humphrey, San Leandro, CA), while the lensometer used was the Lensometer Model 12603 (American Optical Corporation, Buffalo, NY).

4.1.5 Scanning

All scanning was performed by the same investigator (DK). For each scanning session, the subject’s undilated right eye was aligned with the OCT machine. The subject was then asked to gaze at an internal fixation light within the OCT. The machine was programmed to assume a normal, emmetropic eye and to scan along a horizontal line beginning at the temporal edge of the foveal pit and extending nasally 3 mm. Figure 4.1 illustrates the position of the scan on the retina of a study subject. In the figure, the white haze above and below the OCT scan line is an artifact resulting from glare. The OCT was set to image along a line overlapping the fixation point and scanning was begun. The scanning line position was then adjusted so that the image of the foveal pit appeared as deep as possible on the temporal side of the scan. It was assumed that the deepest part of the foveal pit was its center. A sample scan is in Figure 4.2, where one can see the typical placement of the scan relative to the foveal pit.

The procedure used to acquire scans was consistent with the standard clinical used of the OCT. The OCT displays scans continuously at the rate of approximately one per second, and updates the screen image accordingly. Acceptable scans were
4.1.6 Analysis

The acquired scans were exported to an SGI computer workstation for subsequent analysis using custom automatic boundary detection software written for the
Figure 4.2: This figure shows a sample OCT scan with the measurement boundaries marked.

MATLAB software platform as explained in Chapter 7. The software automatically detects the vitreo-retinal junction as the inner retinal boundary and the retinal-choroidal junction as the outer retinal boundary. The retinal thickness was then calculated as the distance between the two boundaries along each A-scan. For the purpose of comparison, the automatic boundary location was manually verified for all scans and corrected when necessary. All inspection and correction was done by the same, unmasked investigator (DK). Separate analyses were performed with both the corrected and the uncorrected boundaries. While the Humphrey 2000 comes with its own automatic boundary detection software, its results are unreliable and cannot be exported for analysis.

A sample measured scan is presented in Figure 4.2 with the marked retinal boundaries highlighted. Figure 4.3 shows the calculated retinal thickness contours for all five scans within a single session (subject 25, session 2). That is, Figure 4.3 shows how the thickness varies across the acquired scan. The fovea is represented by the point of minimum thickness towards the left side of the plot. More variability was found near the fovea, as expected, because the actual retinal thickness changes in this
Figure 4.3: This figure shows the retinal thickness contours from five scans obtained within a single session.

region. Thus small changes in scan placement can cause very different scan profiles in the foveal region.

The retinal thickness value for a scan was calculated for a 1 mm long section located 0.75 mm from the fovea. The mean and standard deviation for each individual subject were calculated over each session (5 scans total per session) and over all sessions (15 scans total). Finally, the mean thickness value was calculated over all subjects using each subject’s mean over 15 scans.

To examine the effects of the within-subject factors scan number (within-session repeatability) and session number (intersession repeatability), a repeated measures analysis was performed using the statistical software package SAS on a PC. For each subject there were 15 observations (3 sessions with 5 scans per session). Separate analyses were performed on the thickness measurements derived from the corrected...
and uncorrected boundaries. The effect of boundary correction was measured by computing the difference between thickness measurements from corrected and uncorrected boundaries for each scan. The mean and standard deviation of these differences were calculated.

To quantify the typical variability occurring between sessions, average thickness values were found for each subject over each individual session and over all three sessions (the latter is assumed to be an estimate of the true value). Then, for each subject, the individual session averages were subtracted from that subject's average over all sessions. These differences indicated how much session averages varied about the true value (assuming that each subject's true retinal thickness was unchanged). The differences in means were then averaged over all patients and all sessions as an indication of normal intersession variability in mean retinal thickness measurements. A similar analysis was performed comparing the retinal thickness measurements derived from individual scans to each subject's average over all scans. This result indicated the normal variation between individual scans.

Two additional statistics were computed for comparison with the literature. For each subject, the standard deviation of thickness values was calculated for each of the three sessions. The within-session standard deviation was then averaged over all subjects for comparison with the value of Hee et al [9]. Each standard deviation was also divided by its corresponding session mean to compute the within-session coefficient of variation (CV). The CVs were then averaged over all subjects and all sessions, for comparison with the results of Baumann et al [2].
Two separate linear regressions were performed to study the effects of pupil size and refractive error on intrasubject variability. The variability for a subject was measured as the standard deviation for that subject's 15 scans. This standard deviation was then regressed against mean pupil size and mean spherical correction.

4.2 Results

For the uncorrected boundaries, the result of the repeated measures analysis testing showed no significant effects for either session number (intersession repeatability) or scan number (within-session repeatability); the p values were 0.529 and 0.509, respectively. Correcting the retinal boundaries yielded similar p values of 0.567 for session number effects and 0.573 for scan number effects. The thickness measurements from corrected and uncorrected boundaries differed by $1.1 \pm 3.3 \, \mu m$ (mean ± standard deviation), which is not significantly different from zero ($p = 0.739$).

The mean and standard deviation for retinal thickness measurements over all patients and all scans was $273 \pm 17$ for uncorrected boundaries and $274 \pm 17 \, \mu m$ for corrected boundaries. The means and standard deviations (SD) of retinal thickness measurements for all patients, as grouped by session, are presented in Table 4.1. As each of the 26 subjects experienced five scans per session, the means are over 130 values. The means and standard deviations of retinal thickness measurements for all patients, as grouped by scan number, are presented in Table 4.2. As three sessions were performed for each of the 26 subjects, the means are calculated over 78 values.

The mean and standard deviation of differences between the individual session averages and the average over all sessions, as calculated for each subject, are presented in Table 4.3. The mean and standard deviation of differences between individual scan...
Table 4.1: This table shows the short-term, between-session repeatability for OCT measurements, by showing the means for the various sessions. Retinal thickness measurements are grouped by session (5 scans per session), showing that the thickness means do not vary between the individual sessions.

<table>
<thead>
<tr>
<th>Session</th>
<th>Number of Subjects</th>
<th>Mean ± SD (µm) from uncorrected boundaries</th>
<th>Mean ± SD (µm) from corrected boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>273 ± 17</td>
<td>274 ± 17</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>273 ± 17</td>
<td>274 ± 18</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>272 ± 17</td>
<td>273 ± 17</td>
</tr>
<tr>
<td></td>
<td>Repeated measures test for intersession effect</td>
<td>p=0.529</td>
<td>p=0.567</td>
</tr>
</tbody>
</table>

averages and each subject’s average are presented in Table 4.3, as well. The mean within-session coefficient of variation was 1.2 ± 0.7 % for corrected boundaries and 1.1 ± 0.8 % for uncorrected boundaries. The mean subject within-session standard deviation is 3.2 ± 2.1 µm for corrected boundaries and 3.0 ± 2.2 µm for uncorrected boundaries.

Pupils ranged from 4.0 - 7.0 mm in the dilated state (mean ± SD = 5.9 ± 0.8 mm). Constricted pupils ranged from 1.5 - 3.5 mm (mean ± SD = 2.6 ± 0.5 mm). The range of refractive errors for the subjects (mean spherical correction) was +1.00 to -10.75 D (mean ± SD = -2.50 ± 2.75). No significant relationship was found by regressing either mean pupil size or mean spherical correction against the standard deviation of each subject’s 15 scans. For the pupil size, the slope of the regression with corrected boundaries was 0.16 (p=0.76); for mean spherical correction, the slope of the regression was -0.03 (p=0.83).
Table 4.2: This table shows the lack of a within-session effect for OCT measurements, by showing that there is no systematic change in thickness value as one repeatedly measures within a session. In the table, retinal thickness measurements are grouped by session (5 scans per session).

<table>
<thead>
<tr>
<th>Scan</th>
<th>Number of Subjects</th>
<th>Mean ± SD (µm) from uncorrected boundaries</th>
<th>Mean ± SD (µm) from corrected boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>273 ± 17</td>
<td>274 ± 17</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>272 ± 17</td>
<td>274 ± 17</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>273 ± 17</td>
<td>273 ± 17</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>273 ± 17</td>
<td>274 ± 17</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>272 ± 17</td>
<td>274 ± 17</td>
</tr>
</tbody>
</table>

Repeated measures test for within-session effect:
p = 0.509  p = 0.573

Table 4.3: This table shows that single measurements tend to cluster close to an individual subject's thickness mean, demonstrating that the measurements are highly repeatable. When session averages are considered, the variability is even less. In the table, the differences between subject thickness averages and thickness measurements from individual scans and sessions are shown.

<table>
<thead>
<tr>
<th></th>
<th>uncorrected boundaries</th>
<th>corrected boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ± SD of session differences averaged over all subjects and all sessions</td>
<td>0.0 ± 2.5 µm</td>
<td>0.0 ± 2.7 µm</td>
</tr>
<tr>
<td>Mean ± SD of individual scan differences averaged over all subjects and all sessions</td>
<td>0.0 ± 4.2 µm</td>
<td>0.0 ± 4.3 µm</td>
</tr>
</tbody>
</table>
4.3 Conclusions

In this study, macular thickness measurements made on undilated subjects using a Humphrey 2000 OCT system were shown to be repeatable within a session and over different independent sessions. The measurements made in this study were performed on scans obtained through the fovea; repeatability for measurements made elsewhere would depend on how easily the scan could be consistently placed at the same location. Intersession repeatability suggests that the OCT might be used to follow patients over time to measure the progression of pathology. Furthermore, the commercially available Humphrey 2000 OCT system, while giving images qualitatively different from those in the literature, gives satisfactory quantitative performance when coupled with the software developed in this dissertation (Chapter 7).

For this chapter, retinal thickness was calculated using both corrected and uncorrected retinal boundaries output by the software. One may ask whether it is necessary to correct the results of the automatic boundary detection algorithm, as this increases the amount of time necessary to obtain retinal thickness information. The corrected averages differed from the original output of the algorithm by $1.1 \pm 3.3 \, \mu m$. This is not significantly different from zero ($p = 0.739$) and will be less than $9.7 \, \mu m$ (99% CI). An advisable strategy might therefore be to view the algorithm's output to verify that there are no gross errors, but to otherwise accept its results. Software allowing correction of the boundaries could be a useful addition to the OCT. Because the difference is not clinically significant, only the corrected results are discussed below.

The average retinal thickness of $274 \pm 17 \, \mu m$ over all 26 subjects using the corrected bounds. Our subject population was generally young and with the exception of four subjects, the refractive errors ranged from +1.00 D to -5.00 D. Our average
thicknesses are also consistent with the values presented by Hee, et al[9]. They measured the retinal thickness within a 3 mm radius disk surrounding the fovea by interpolating the results of six foveal scans placed 60 degrees apart. All scans were acquired in one session. The average foveal thickness for all subjects was reported to be 174 ± 18 μm. Along a horizontal line segment extending nasally from the fovea, they reported the average in 73 subjects to be 260 ± 16 μm for the segment extending 0.5 to 1.5 mm from the fovea, and 255 ± 16 μm for the segment extending 1.5 to 3.0 mm from the fovea.

The technique of Hee, et al. provided six foveal thickness measurements for each subject, all within the same session. The within-session standard deviation within the six measurements of each subject was reported to be distributed with a mean of 11 and a standard deviation of 6 μm. In this study, each session yielded five thickness measurements; a similar calculation as by Hee et al. for the purposes of direct comparison yields an average subject within-session standard deviation of 3.2 ± 2.1 μm with the corrected boundaries.

For the current study, the data in Table 4.3 show that the differences between session averages and total patient averages are distributed with a standard deviation of 2.7 μm. Similarly, the differences between individual scans and subject averages are distributed with a standard deviation of 4.2 μm. These standard deviations best quantify the variability within this study, and are of the same order of magnitude as in Hee et al. The larger size of their variability assessment may result from their measurement of thickness at a single point rather than within a region.

The data from this study may be better interpreted by using confidence intervals. There is a 99% confidence that session averages will be within 7.0 μm of the true
subject value using corrected boundaries. Similarly, there is a 99% confidence interval that individual scan averages will be within 11.2 \( \mu m \) using corrected boundaries. Only a small decrease in the variability is achieved by averaging five scans per session. Hence, it may be reasonable for a clinician to accept retinal measurements from only one or two scans.

The within-session reproducibility results found here can also be compared to those of Baumann, et al [2]. They obtained six vertical scans of length 2.88 mm through the fovea of 18 eyes. The scans were divided into seven sections and the sections were treated individually. The coefficient of variation (standard deviation divided by mean) for each scan segment was averaged over all subjects, and these averages were presented for both manual and automated retinal boundary determinations. The segments of their scans most relevant to this study were the most superior and inferior segments (1.12 to 1.44 mm from the fovea). For these they found mean coefficients of variation (CV) of 4.1% and 3.8%, respectively. The average CV of 1.2% achieved in the current study is comparable.

The ability to perform OCT measurements in undilated subjects extends the potential uses of this instrument. While Baumann, et al. obtained their single session OCT measurements in undilated patients, the subjects’ pupil sizes were not reported. It is not uncommon for circumstances to make dilation either difficult or undesirable. Adequate dilation may be difficult in patients having extended topical pilocarpine therapy, exfoliation syndrome[77], pseudoexfoliation syndrome, chronic diabetes, and, occasionally, old age. Subject variability does not seem to be related to pupil size, though the smallest pupil diameter for which macular thickness measurements can be made routinely with OCT was not determined in this study. The use of OCT in
undilated patients presents further limitations. In particular, the fundus is frequently
difficult to visualize while scans are being obtained. Thus the patient's cooperation
in fixation or the presence of clearly identifiable OCT landmarks such as the foveal
pit are necessary for scans to be located precisely on the fundus. Scans around the
optic nerve head would most likely be compromised in undilated patients for these
reasons.

The OCT measurements made in this study were made without knowledge of
subject axial length or refraction. Because of the design of the OCT optics, these
parameters are necessary for distances in the transverse direction (i.e. along the
direction of the scan line) to be measured accurately [9]. A standard, emmetropic eye
(plano, 23 mm axial length) was assumed for all subjects, which likely caused some
intersubject variations from the assumed scan length. However, these errors would
remain constant for each subject from session to session, and therefore wouldn't be
expected to affect repeatability. Moreover, measurements in the axial direction (i.e.
into the retina) do not depend on the refraction or axial length of the eye. Thus
thickness values are not affected. To date, no study has been published where axial
length and refraction were used to ensure accurate scan length measurements.

In conclusion, the OCT offers a way to make measurements of retinal thickness
in patients that are repeatable over different sessions. In particular, the Humphrey
2000 OCT system gives repeatability comparable to that of the prototype systems in
the literature. The automatic boundary marking system used here can be expected
to give repeatable measurements close to those marked manually. Moreover, this
intersession repeatability has been demonstrated in undilated patients, extending the
OCTs noninvasiveness.
CHAPTER 5

LONG TERM VARIABILITY

5.1 Materials and Methods

5.1.1 Equipment

Again, the Humphrey 2000 OCT system (Zeiss-Humphrey, San Leandro, CA) was used in this study. It is a commercial implementation of the prototype systems that have been previously reported [2, 10, 22, 75]. The Humphrey 2000 OCT uses an 850 nm light source with a patient exposure range of 200 μW to 750 μW. Each Humphrey 2000 image is comprised of 100 A-scans, with 500 pixels per A-scan.

5.1.2 Subjects

Eighteen normal volunteers were recruited for this study. The volunteers were selected from research associates and clinic staff, and informed consent was obtained. The subjects included 10 men and 8 women, and ranged in age from 20-52 years (mean ± SD, 30 ± 9 years). For each subject, the mean spherical correction was found from automatic refractometer (model 585; Allergan Humphrey) or lensometer (model 12603; American Optical, Buffalo, NY) to be -2.55 ± 2.90 D. The range of corrections was +0.13 D to [minus] 10.75 D. Exclusion criteria included existing
or previous known retinal disease. The protocol was approved by The Ohio State University Biomedical Sciences Institutional Review Board.

5.1.3 Scanning

For this study, each subject completed three scanning sessions. The first two sessions were completed within one hour, while the third session occurred after a separation of 15.5 ± 3.6 (range 12-20) months. Thus the first two sessions are referred to as the initial “visit” and the last session as the second visit. All scanning was performed by the same investigator (DK), and all sessions were completed with identical protocols. For each session, five horizontal scans of nominal length 3 mm were obtained through the fovea of the subject’s undilated right eye. Thus each subject had a total of 15 scans. Figures 5.1 and 5.2 show typical scan placement and scan appearance, as captured from the Humphrey system. Internal fixation was used to locate the scan line, and the scan line was positioned so that the image of the foveal pit appeared as deep as possible.

5.1.4 Measurements and Analysis

The OCT image data files were exported to an SGI computer (Silicon Graphics, Mountain View, CA) for analysis. The data files contain the actual signal intensities recorded by the OCT machine, and their use avoids potential quantization artifacts generated by the pseudocolor, onscreen display. The analysis was performed using custom software, written for the MATLAB software platform (The Mathworks, Natick, MA), that automatically detects the vitreo-retinal junction as the inner retinal boundary and the retinal-choroidal junction as the outer retinal boundary [78] as described in Chapter 7. The retinal thickness along the length of the OCT B-scan was
Figure 5.1: A sample fundus view is shown with the scan placement illustrated as a bright, horizontal line located nasally to the fovea.

Figure 5.2: A typical OCT scan is displayed in grayscale, with various anatomical features marked.
Figure 5.3: This sample OCT scan has the retinal boundaries, as found with the custom software, outlined in white, and the regions measured for the thickness analysis have been shaded with gray rectangles. In the figure, the black vertical line denotes the foveal center.

then measured as the distance between each boundary, using the known scale factor of 4 mm per image pixel [10]. Examples of the automatically detected boundaries can be seen in Figure 5.3, where they are outlined in white. The automatically determined retinal boundaries were manually inspected and corrected where necessary; this was done in a random order by a masked investigator (DK) to ensure that there was no bias.

For each scan, the average retinal thickness was measured over a 0.06 mm region centered on the fovea and over a 1 mm region located 0.75 mm nasal to the fovea. The 0.06 mm regional average is the average retinal thickness for the A-scan at the foveal center and its neighbor to either side. This averaging helps lessen the noise that is rampant in OCT images. Figure 5.3 illustrates these regions for a sample scan. The
mean and standard deviation of both the retinal and foveal thickness measurements were found over all scans obtained by all subjects.

To examine the effect of the within-subject factor, session number, a repeated-measures analysis was performed using a statistical software package (SAS, Cary, NC). The sessions were separated over time, and so this analysis would detect any consistent effects caused by the long term separation. To characterize the measurement variability, each subject had the mean retinal thickness calculated for each session; these session averages were thus over five scans. The difference between the first and third session averages represents the long term change in retinal thickness between the two visits, and this change was averaged over all subjects. This average represents long term variability, and was compared to the average difference between the first and second sessions, which represents short term variability. The comparison was done using a two-sample F-test to first determine whether the variances of the differences were equal, and then using the appropriate two-sample t-test. For all statistical tests, a p value less than 0.05 was considered significant.

For additional measures of the variability between the two visits, an intersession coefficient of variation (CV) was calculated for each subject as the standard deviation of the thickness values from that subject's 15 scans divided by that subject's mean thickness value over all 15 scans. The within-session CV was calculated as the standard deviation of the five thickness values obtained within a single session divided by the mean retinal thickness for that session. Each subject then had one measure of intersession CV and three measures of within-session CV. These CV values were averaged over all subjects. The intersession and within-session standard deviation values used to calculate the CVs were also recorded and averaged over all subjects.
5.2 Results

For the macular thickness measurement over a 1 mm region, the repeated measures analysis found no significant effect (p = 0.523) for the time separation. The same was true for the foveal thickness measurements (p=0.810). Figures 5.4A and 5.4B show the retinal thickness curves from two different subjects; these curves plot retinal thickness over the entire 3 mm scan length. The curves from the five scans obtained in the first session (initial visit) are shown as dashed lines, and the curves from the five scans obtained in the third session (second visit) are shown as solid lines.

For the retinal thickness measurement over the 1 mm region, the average over our subject population was 271 ±17 μm (222 - 320, 99%CI). The mean within-session and intersession standard deviation and CV are reported in the table. The mean difference between the first and third session averages, which reflects the variability between visits (long term), was 0.9 ± 3.9 μm (-10.4 - 12.2, 99%CI), while the mean difference between the first and second sessions, which reflects the variability within a single visit (short term), was 0.6 ± 4.2 μm (-11.6 - 12.8, 99%CI). A two-sample F-test for variances found no difference between the variances of the two sets (p=0.827). A two-sample, equal variances t-test on the two sets found their means to be equal as well (p=0.833).

For the foveal thickness measurements, the average over our subject population was 157 ± 21 μm (96-218, 99%CI). The mean within-session and intersession standard deviation and CV are reported in the table. The mean difference in measured thickness between sessions one and three was 1.3 ± 8.4 μm (-23 - 26, 99%CI), and the mean difference between sessions one and two was 0.4 ± 7.7 μm (-22 - 23, 99%CI). A two-sample F-test for variances found no difference between the variances of the
Figure 5.4: A and B: These figures illustrate, for two different subjects, the overall similarity of the retinal thickness curves generated from the two different visits. In each figure, the dashed lines denote the five scans obtained at the first session of the initial visit, and the solid lines denote the five scans obtained during the second visit. The individual plot lines have been shifted horizontally to align the foveal center.

two sets (p=0.719). A two-sample, equal variances t-test on the two sets found their means to be equal as well (p=0.747).

5.3 Conclusions

Macular thickness measurements made on normal subjects using a Humphrey 2000 OCT system are repeatable between sessions separated by 15 months, and the variability over 15 months is identical to that over one hour. This is true for thickness measurements over a 1 mm portion of a 3 mm OCT scan as well as over a much smaller, 0.06 mm perifoveal region. These results suggest that no thickness variations detectable by OCT occur over these time periods, and thus the short term variability
Table 5.1: This table illustrates the similarity of the long and short term variability measures by presenting the intersession and within-session standard deviation (SD) and coefficient of variation (CV) for the two retinal thickness measurements used in this study. Because the three sessions span two visits separated by an average of 15 months, the intersession values reflect long term variability while the within-session values reflect short term variability.

<table>
<thead>
<tr>
<th></th>
<th>1 mm retinal region</th>
<th>Fovea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-session SD</td>
<td>3.5 ± 1.6 μm</td>
<td>5.5 ± 2.2%</td>
</tr>
<tr>
<td>Intersession SD</td>
<td>4.1 ± 1.2 μm</td>
<td>9.4 ± 2.0 μm</td>
</tr>
<tr>
<td>Within-session CV</td>
<td>1.3 ± 0.6%</td>
<td>1.5 ± 0.4%</td>
</tr>
<tr>
<td>Intersession CV</td>
<td>1.5 ± 0.4%</td>
<td>6.1 ± 1.3%</td>
</tr>
</tbody>
</table>

reported here and elsewhere in the literature may be used for patient follow up schedules of this duration. These results confirm for the 10 μm precision of OCT what has been presumed to be true based on appearance with slit lamp biomicroscopy.

The difference between the session averages, when interpreted in terms of confidence intervals (CI), indicates the magnitude of the variability. For the foveal thickness measurement, one expects (99% CI) that two measurements obtained within one hour will differ by less than 23 μm and that two measurements separated by 15 months will differ by 26 μm. Similarly, for the 1 mm retinal region, one has short and long term expected differences (99% CI) of less than 12.8 μm and 12.2 μm, respectively. A qualitative comparison of the long and short term variabilities may also be performed by comparing the within-session and intersession values of both
the standard deviation and the CV. In both cases the intersession and within-session values appear very similar, indicating that a time separation of 15 months does not substantially increase the variability of thickness measurements.

The mean retinal and foveal thickness values found here are comparable to those found by Hee, et al. [10]. They found average foveal thickness in 73 subjects to be $174 \pm 18 \mu m$. Along a horizontal line extending nasally from the fovea, they found the average retinal thickness to be $260 \pm 16 \mu m$ for the region 0.5 to 1.5 mm from the fovea and $255 \pm 16 \mu m$ for the region 1.5 to 3.0 mm from the fovea. The 15 month variability measures in this study may also be compared with the short term variability measures in the literature. Hee et al. [10] obtained six scans through the fovea at one sitting. Their standard deviation for the six foveal thickness measurements was $11 \pm 6 \mu m$, which is actually larger than the intersession and within-session values found here. However, they measured foveal thickness at a single point, which implies measuring an individual A-scan at the foveal center. This study defined foveal thickness as the average of three A-scans centered on the fovea; this averaging would be expected to remove some of the variability. Finally, Baumann et al. [2] calculated CVs of 4.1% and 8% for retinal regions comparable to the 1 mm segment and peri-foveal regions used here. Thus the long term variability found in this study is comparable to, or even smaller than, the short term variability that has been reported.

The repeatability and variability reported here were demonstrated with scans obtained through the fovea of normal subjects. Pathologies will naturally have their own time variation, and monitoring thickness elsewhere on the retina imposes the challenge of consistently locating the OCT scan line in the same location. Retinal
thickness is known to decrease with age in rats and is suspected to decrease in humans due to ganglion cell death. The lack of any significant effect detected by the repeated measures analysis suggests that any such changes are too gradual to be detected over 15 months for the subject age range investigated here. Finally, the variability of the foveal thickness measurement is higher than that for the 1 mm retinal region. One cause is that the foveal thickness measurement is the average from three A-scans, as opposed to 34 A-scans for the 1 mm regional thickness; greater averaging lessens the effect of noise. Another cause, however, is that the fovea is a small pit on the retinal surface. Thus even slight displacements of the scan line from the foveal center result in different measured thickness values.

In conclusion, macular thickness measurements made with OCT have been shown to be repeatable and stable over time frames greater than one year, with variability identical to that of short term measurements. Thus the assumption of retinal stability in normal subjects, which was previously based on slit lamp biomicroscopy, has been shown to be valid to the 10 mm precision of OCT. The expected amount of variability is less than 12 \( \mu m \) for the 1.0 mm macular region and 26 \( \mu m \) for the foveal measurements (99% confidence); these represent changes of 5% and 17% of normal thickness values for the macula and fovea, respectively. Thus, OCT offers a sensitive, precise, and objective method to monitor changes in retinal thickness.
CHAPTER 6

EFFECTS OF SLD POWER AND PUPIL DILATION

6.1 Material and Methods

6.1.1 Instruments

The OCT used in this study is still the commercially available Humphrey 2000 OCT system (Zeiss-Humphrey, San Leandro, CA). It still uses an 850 nm superluminescent diode, whose power can be varied from 200 to 750 $\mu$W. Each Humphrey 2000 image is still composed of 100 A-scans, with 500 pixels per A-scan and a manufacturer reported, longitudinal resolution of 10 $\mu$m.

6.1.2 Subject selection

Twenty one normal volunteers were recruited from research associates and clinic staff, and informed consent was obtained. The subjects included 12 men and nine women, who ranged in age from 25-58 years (median 35 years). Refractive error for each subject was recorded from best-corrected visual acuity and either manifest refraction using automatic refractometer (model 585; Allergan Humphrey) or refractive prescription. The mean spherical correction was -2.79 ± 2.72 D. The study protocol was approved by The Ohio State University Biomedical Sciences Institutional Review Board and followed the tenets of the Declaration of Helsinki.
6.1.3 Scan acquisition

For this study, each subject completed four scanning sessions. The sessions were designed to explore all four possible concomitant variations of the two OCT parameters that were investigated: SLD power (500 $\mu W$ vs. 750 $\mu W$) and subject dilation state (dilated vs. undilated). Thus when initially undilated, each subject underwent one scanning session with a 500 $\mu W$ SLD power setting and one with a 750 $\mu W$ setting. Each subject was then dilated with 1% mydriacyl and 2.5% tropicamide and scanned for two more sessions at 500 and 750 $\mu W$. The same eye was used for all scanning sessions (14 OD and 7 OS, as per individual subject preference), and each subject had the same scan operator complete all four scanning sessions.

Other than the SLD power and subject dilation, each session followed an identical scanning protocol. For each session, five horizontal scans of nominal length 3 $mm$ were obtained through the foveal center. Scans were aligned using the OCT’s internal “landmark” light as a fixation target for the foveal center. Once scanning was initiated, the line position was adjusted so that the foveal pit appeared as deep as possible.

6.1.4 Data analysis

The raw, binary data files for the B-scans were then exported to a computer workstation for analysis using custom software written in MATLAB [78] as stated in Chapter 7. This software automatically locates the retinal boundaries within an OCT B-scan and determines the retinal thickness as the distance between these boundaries. Using this software, retinal thickness was measured for each scan over a 1 $mm$ region located 0.75 $mm$ from the fovea and for a 0.06 $mm$ region centered on the fovea. The
0.06 mm region corresponded to the A-scan at the foveal center and one A-scan to either side; this region was chosen rather than a single A-scan to ameliorate the effect of OCT image noise. Figure 6.1 shows a sample OCT scan with the marked retinal boundaries shown in white and the measured retinal regions shaded in gray. Each automatic determination of the retinal interfaces was manually verified and corrected when necessary; to avoid bias, the scans were inspected in a random order with the scan identity masked. The verification was performed using the grayscale images to avoid the artifacts color can introduce. Identical analyses were performed for the thickness values from the 1 mm retinal region and the foveal region.

To analyze the within-subject factors SLD power and pupil dilation state, a repeated measures analysis was performed using the statistical software SAS (Cary,
NC). This analysis determined whether systematic differences in measured retinal thickness were caused by these scanning parameters. To further investigate the effect of the scanning parameters, each subject had their mean retinal thickness calculated over all four sessions, as well as their mean thickness for each individual session (5 scans per session). Each session average thus represented a subject’s thickness value obtained with one set of scanning parameters. For each subject, then, the differences were calculated between their average over all sessions and each of their individual session averages. A scatter plot of these differences, grouped by session, allowed a qualitative view of the effect of scanning parameters on thickness measurements.

To quantify these effects, the session with dilated pupils and SLD power of 750 $\mu W$ was arbitrarily chosen as a base standard. For ease of discussion, this will be referred to as the Dilated-750 $\mu W$ session, and similar notation will be used for the other sessions. The effect of pupil dilation was then calculated for each subject as the difference between this base session and the session with undilated pupils and SLD power of 750 $\mu W$ (Undilated-750 $\mu W$). These differences were then averaged over all subjects. Similarly, the effect of the SLD power was calculated as the difference between the base session and the session with dilated pupils and SLD power of 500 $\mu W$ (Dilated-500 $\mu W$).

Finally, a regression analysis was used to investigate the effect of the B-scan signal intensity on the measured retinal thickness. The retinal thickness values from the 21 subjects could not be directly pooled for analysis because of individual differences in retinal thickness. Instead, relative values for both the scan intensity and the measured thickness were used to remove the between-subject differences, using the following procedure.
First, for each subject, the scan intensity for each of their 20 B-scans was defined as the mean signal intensity of the pixels contained within the retinal boundaries of the 1 mm retinal measurement region. This region was chosen because it is the actual portion of the scan that was measured for thickness. Furthermore, the intensity values were taken from the Humphrey’s binary data files; these files do not contain units and so values in this paper will be labeled with intensity units. The mean scan intensity for each subject was defined as the average scan intensity of that subject’s 20 B-scans. Finally, for each subject, the relative scan intensity for each of their 20 B-scans was defined as the difference between the scan intensity for that B-scan and the subject’s mean scan intensity. Similarly, the relative thickness value for each scan was the measured retinal thickness value for that scan minus the mean retinal thickness value for the corresponding subject. The relative thickness values of all 420 B-scans collected in this study were then regressed against their relative scan intensities to determine if a relationship existed.

6.2 Results

For the macular thickness measurement over a 1 mm region, the repeated measures analysis found no significant effect for the SLD power (p = 0.514), pupil dilation (p=0.109), or the interaction of SLD power with pupil dilation (p=0.359). Similarly, for the foveal measurements, no significant effects were found for SLD power (p=0.714), pupil dilation (p=0.643), or the interaction of SLD power with pupil dilation (p=0.295).

The mean retinal thickness over all subjects and all scans was 275 ± 16 mm for the 1 mm retinal region and 160 ± 21 mm for the foveal region. Figure 6.2 shows
Table 6.1: Reported above are the average, between-session differences in thickness measurements for this study. The session conducted with dilated pupils and 750 μW SLD power setting has been arbitrarily taken as a standard to compare sessions in which the scanning power is varied or the pupils are undilated.

scatter plots of the differences between each subject's individual session averages and total average over all scans, for the 1.0 mm retinal region and the foveal region. Table 6.2 shows the average measured thickness difference between the Dilated-750 μW and Undilated-750 μW sessions and between the Dilated-750 μW and Dilated-500 μW sessions.

The regression of the relative thickness value versus the relative scan intensity was not significant for either the 1 mm retinal region (slope ± SEM = 0.0013 ± 0.0016 mm/(intensity unit), p = 0.407, R² = 0.00165) or the foveal region (slope ± SEM = 0.0060 ± 0.0031 mm/(intensity unit), p=0.052, R² = 0.00897), though the foveal value was marginally insignificant. Scatter plots of these relative thickness values versus relative scan intensity are shown in Figure 6.1.

6.3 Conclusions

Neither pupil dilation nor SLD power variations from 500 to 750 μW significantly affected the measured retinal or foveal thickness, using the techniques in this study. Thus patient data acquired within this range of scanning conditions may be directly

<table>
<thead>
<tr>
<th>Dilated-750 μW vs. Undilated-750 μW (mean ± SD)</th>
<th>Dilated-750 μW vs.Dilated-500 μW (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mm retinal region 0.5 ± 3.1 μm</td>
<td>-0.1 ± 3.1 μm</td>
</tr>
<tr>
<td>foveal region 0.8 ± 6.8 μm</td>
<td>1.3 ± 5.2 μm</td>
</tr>
</tbody>
</table>
Figure 6.2: These scatter plots show that varying the scanning parameters has no consistent effect on the thickness measurements (fovea in A and 1 mm retinal region in B), as illustrated by the fact that the stars in each column are all centered about zero. The four columns in each plot represent one set of scanning parameters, and each star represents the difference between a subject’s measurement under that set of parameters and their average measurement over all sets of parameters. The four sessions are labeled on the horizontal axis; “u500” denotes the Undilated-500 $\mu W$ session, and likewise for the other three sessions.

Figure 6.3: These scatter plots demonstrate the lack of effect of B-scan signal intensity on measured retinal thickness. The axes are for relative values of intensity and thickness to remove the effect of between-subject thickness differences.
compared. Pupil dilation and SLD power affect OCT scan appearance by affecting the interference signal intensity within the scan. However, the measured thickness for each scan did not significantly depend on its signal intensity. Though the regression for the relative foveal thickness is near statistical significance at $p=0.052$, it is far from clinical significance with a calculated slope of $0.0060 \pm 0.0031 \, \mu m / (\text{intensity unit})$.

The clinical insignificance can be deduced from Figure 6.1, which suggests that for the range of intensities seen in this study, the regressed slope would imply a measured foveal thickness difference of $4.8 \, \mu m$; this is less than the normal variability for OCT derived thickness measures [11]. Moreover, the large amount of scatter, as reflected by the low $R^2$ values, suggests that any relationship between measured thickness and scan intensity is tenuous. Finally, in Figure 6.1, the amount of vertical scatter does not appear to vary with relative signal intensity. This suggests that the variability in thickness measurements, i.e. the degree of deviation from a subject's thickness mean, does not change with signal intensity.

This lack of an effect is surprising, given the difference in appearance between the high and low intensity B-scans in Figures 2.1 and 3.7. Even with the normalized color and grayscale maps, the retina appears dimmer in the low intensity scans, with larger regions that are close to the background black of the vitreous humor. Moreover, the amplitude of the speckle noise appears to be larger in the high intensity image, filling in more of the otherwise black photoreceptor layer. Furthermore, the A-scan signal peaks may be broader in the high intensity image. The overall effect is that the high intensity scans may be perceived to be thicker, particularly when displayed using the Humphrey absolute color map. The absolute color map is particularly important because this is what the OCT operator sees when scans are acquired.
Despite these apparent differences, however, the retinal boundaries remain distinct, and their separation within the scan is not changed in any consistent, significant way. In Figure 2.1, for example, the higher intensity scan is 6.3 $\mu$m thinner, and in Figure 3.7, it is 5.7 $\mu$m thicker. The scatter plots in Figure 6.2, illustrating the differences between the individual session averages and the subject means, are also consistent with a lack of any significant effect from changing the scanning parameters. The differences are approximately centered on zero for all four sessions, indicating that varying the scanning parameters causes no consistent, systematic changes in measured thickness. This is supported by the confidence intervals for the intersession thickness differences for the 1 mm retinal region. The expected difference (99% CI) between the Dilated-750$\mu$W and Undilated-750$\mu$W sessions is less than 9.2 $\mu$m, and the expected difference (99% CI) between the Dilated-750$\mu$W and Dilated-500$\mu$W sessions is less than 8.8 $\mu$m. These difference values are similar to the 7 $\mu$m expected variation reported for sessions conducted with identical scanning parameters [34].

This lack of an effect is important for compiling a comprehensive picture of OCT from the literature, where results are reported under a variety of dilation states and scanning powers. Moreover, the lack of an effect from pupil dilation is particularly important. As OCT evolves from a research tool to a common, clinical tool, confidence in retinal thickness measurements made through undilated pupils becomes important. Patients may wish to avoid dilation or may present with poorly dilated pupils that make funduscopy difficult. The mean retinal thickness values measured here are consistent with those previously reported [9, 11].

One must note that the thickness measurements in this study were made using the raw, binary data files for each scan and a custom boundary detection system allowing
manual verification and correction of obvious errors. Finally, one must also note that extremely weak scans, whose brightest regions are displayed by the Humphrey as blue with green highlights, were not acquired for this study and so conclusions cannot be drawn for them.

Interestingly, the scan signal intensity was not tightly related to the scanning parameters. The high and low intensity scans in Figures 2.1 and 3.7 were acquired as part of the Dilated-750μW and Undilated-500μW sessions, respectively, but this was not necessarily the rule. Scans acquired with an SLD power level of 750 μW generally had higher signal intensities than those acquired with 500 μW. However, for a given SLD power level, similar scan intensities could be achieved for both dilated and undilated pupils, so long as the pupils were properly aligned. Similarly, pupil misalignment caused very low intensity scans with either SLD power setting. Thus the SLD power level sets a ceiling for the maximum signal intensity that can be attained, but pupil misalignment may degrade the scan intensity from this ceiling.

The signal intensity within B-scans has also been used to study the nerve fiber layer [79]. In that study, a normalized grayscale was used similar to that used here, but the values were scaled between 0 and 1. However, the normalized value at each pixel was reported as a relative reflectivity value. Care must be taken to note that the values in the raw data files and the colors in the Humphrey display denote the interference signal intensity, from which the retinal reflectivity is inferred. Reflectivity is an intrinsic tissue property. The retinas in the high and low intensity scan pairs shown in Figures 2.1 and 3.7 have the same reflectivity even though the scans appear different.
In conclusion, OCT measurements of retinal thickness are repeatable over differences in pupil dilation and SLD power setting. Thus the results of different studies using different scanning parameters may be directly compared. Moreover, OCT scans with low signal intensities, of the ranges used in this study, yield reliable retinal thickness measurements. Hence if a high intensity scan cannot be obtained for a patient, a lower intensity scan may be used with confidence to measure retinal thickness. This finding may increase the clinical utility of OCT.
CHAPTER 7

RETINAL BOUNDARY DETECTION

7.1 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. Patient motion frequently causes artifacts ranging from undulations to apparent breaks in the retinal image, examples of which are in Figures 7.1 through 7.3. The Humphrey 2000 requires one second to acquire all 100 A-scans, which is sufficient time for involuntary eye motion to occur. The OCT technician may save such scans, and so our boundary detector must be able to cope with them despite the assumption of retinal smoothness. However, images featuring large dislocations are difficult even for humans to interpret, and so they are routinely rejected and need not be considered by our algorithm.

Because, the OCT acquires each A-scan individually, we apply one dimensional edge detection to each image column individually, similar to Thune et al. [52]. In fact, the rampant speckle noise and breaks in the images make the use of two dimensional
Figure 7.1: A sample OCT scan showing the undulations which may be present. These undulations may be the result of tiny, involuntary eye motions.

Figure 7.2: Another sample OCT scan showing undulations. These larger undulations can result from eye motion or motion of the patient's head. This scan has a small break near column 90.
edge detectors problematic. Indeed, the spatial coherence assumptions used in the design of most two dimensional, 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, which can confound the output of two-dimensional kernels. Finally, each A-scan penetrates both retinal boundaries, and so we assume that every image column intersects exactly two boundaries.

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.

Figure 7.3: Another sample OCT scan showing a break in the scan. Scans featuring breaks larger than this would, most likely, not be saved by the OCT technician.
To select, organize, and interpret the detected edges, we developed a mathematical model for normal boundary contours. We defined a discrete coordinate system, labeled in Fig. (3.3), 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 OCT images.

We can model the boundary displacement between adjacent scans as a \( M^{th} \) 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

\[
b(n) - b(n - 1) = \sum_{i=1}^{M-1} a_i (b(n - i) - b(n - i - 1)) + N(n) \tag{7.1}
\]

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 \( M^{th} \) order. Thus we are using an autoregressive mathematical model for this Markov process. Moreover, \( N(n) \) in Eq. (7.1) represents the prediction errors as random noise distributed identically and independently over \( c \). Finally, also note
that Eq. (7.1) 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

\[
b(n) - b(n + 1) = \sum_{i=1}^{M-1} a_i (b(n + i) - b(n + i + 1)) + N(n) \tag{7.2}
\]

Note that the two sets of \( a_i \) in Eq. (7.1) and (7.2) 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 estimate for the \( a_i \) in the Markov model can be found as described by [15]. We will first build the discussion for Eq. (7.1); the discussion for Eq. (7.2) is very similar. Thus, define the displacement, \( d \), between boundary locations in adjacent A-scans as

\[
d(i) = b(i) - b(i - 1) \tag{7.3}
\]

Eq. (7.1) predicts \( d(n) \) from \( d(n-1), \ldots, d(n-M+1) \), and so using the generic labels of \( x \) and \( y \) for independent and dependent variables, respectively, and noting that \( x \) is a vector,

\[
x(n) = \begin{bmatrix} d(n-1) & d(n-2) & d(n-3) & d(n-4) \end{bmatrix} \\
y(n) = d(n) \tag{7.4}
\]

As detailed in [15], the cross correlation vector \( k_{xy} \) between \( d(n) \) and \( \{d(n-1), \ldots, d(n-M+1)\} \) and the autocorrelation matrix, \( K_{xx} \), for \( \{d(n-1), \ldots, d(n-M+1)\} \) are
\[ k_{xy} = E\left[ d(n) \, d(n - 1) \cdots d(n) \, d(n - M + 1) \right]^T \] (7.5)

\[ K_{xx} = E\left[ \begin{array}{cccc} d(n - 1) & d(n - 1) & \cdots & d(n - 1) \\ \vdots & \ddots & \vdots \\ d(n - M + 1) & d(n - 1) & \cdots & d(n - M + 1) \end{array} \right] \] (7.6)

Here \( E \) indicates that the individual elements in \( k_{xy} \) and \( K_{xx} \) are expected values.

Continuing from [15], the minimum mean squared error estimate for the vector \( a = [a_i] \) is

\[ a = k_{xy}^T K_{xx}^{-1} \] (7.7)

Finally, to predict \( a_i \) for Eq. (7.2), we would use similar \( k_{xy} \) and \( K_{xx} \) but redefine the displacement as

\[ d(i) = b(i) - b(i + 1) \] (7.8)

and the independent and dependent variables as

\[ x(n) = \begin{bmatrix} d(n + 1) & d(n + 2) & d(n + 3) & d(n + 4) \end{bmatrix} \]

\[ y(n) = d(n) \] (7.9)

One option for forming the training set to use Eq. (7.7) would have been to manually mark the boundaries in a set of images. However, to avoid this onerous task, initial \( a_i \) and \( M \) were chosen and an iterative process was started which converged to more optimal values. Thus, we initially chose \( M = 5 \) and

\[ [a_i] = \begin{bmatrix} 0.2757 & 0.1654 & 0.0993 & 0.0596 \end{bmatrix} \] (7.10)
This initial \([a_i]\) was a weighted sum, where each term was 0.6 times the previous term and the weights were scaled to sum to 0.6; we set these values based on intuition and some trial and error. We then implemented the rest of our algorithm, as described in Section 7.2, 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 different individuals [34, 80]. 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.

Thus, from each of the training images, the displacement between each adjacent A-scan pair was found, according to Eqs. (7.3) and (7.8). We only used the boundary locations in 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 \((x, y)\) pairs for each boundary, according to Eq. (7.9) for the \([a_i]\) in Eq. (7.1), and according to Eq. (7.10) for the \([a_i]\) in Eq. (7.2). Then, the \((x, y)\) pairs from all of the 330 training scans were grouped and least squares fits to Eqs. (7.1) and (7.2) were performed for the upper and lower boundaries.

In particular, let \(d_i(n)\) represent the difference between the \(n^{th}\) A-scan and its neighbor (left or right depending on context) for the \(i^{th}\) OCT image in our training set of 330 images. We will again focus on the development for Eq. (7.1) first. We thus define

\[
x_i(n) = \begin{bmatrix} d_i(n-1) & d_i(n-2) & d_i(n-3) & d_i(n-4) \end{bmatrix}
\]

\[
y_i(n) = d_i(n)
\]

The data for the least squares fit is then collected from the 330 training images by forming the matrices
\[
X = \begin{bmatrix}
    x_1(10) \\
    \vdots \\
    x_1(94) \\
    \vdots \\
    x_{330}(10) \\
    \vdots \\
    x_{330}(94)
\end{bmatrix} \quad Y = \begin{bmatrix}
    y_1(10) \\
    \vdots \\
    y_1(94) \\
    \vdots \\
    y_{330}(10) \\
    \vdots \\
    y_{330}(94)
\end{bmatrix}
\]

(7.12)

and the \([a_i]\) are calculated as

\[
a = \left( X^T X \right)^{-1} X^T Y
\]

(7.13)

Note that in Eq. (7.13), the effect of the matrix multiplications and inversions is to create the expected value estimates for Eq. (7.7) and perform the necessary multiplications. Also note that these displacements, \(d(n)\), were calculated using the boundaries as found before Step 6, in Section 7.2. 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 7.1, and are in the form mean ± standard error of the mean (SEM). Note that writing Eq. (7.1) and (7.2) in terms of displacements rather than the actual boundary positions improved the regression by removing a sizable, consistent DC value.
Table 7.1: This table has the values for the boundary detector’s Markov model coefficients. The $a_i$ are shown as estimated by our least squares regression.

Using the Pearson statistical criterion $p < 0.05$, all of the $a_i$ in Table 7.1 are significant, and statistically significant differences exist between the $a_i$ for Eq. (7.1) and (7.2). 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 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 5 (resulting, therefore, in 4 difference values), the performance evaluation that is reported in the Results section was also performed using models of orders 3 through 9. Over this range, the performance was very similar, suggesting that this modeling technique is not sensitive to order number. However, models of order 4 and higher exhibited less error than the third.
order model, and so a model order of 5 was considered a good compromise between performance and model complexity.
**Outline of Approach**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input image and filter twice with 4x4 median filter to reduce speckle noise</td>
</tr>
<tr>
<td>2</td>
<td>Filter each column with the second derivative of Gaussian, acting as a one dimensional edge detection kernel</td>
</tr>
<tr>
<td>3</td>
<td>Choose the strongest two edges, scaled by relative contrast, as the initial approximation to the inner and outer boundaries.</td>
</tr>
<tr>
<td>4</td>
<td>Segment the initial approximations at boundary discontinuities</td>
</tr>
<tr>
<td>5</td>
<td>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 &quot;determined.&quot;</td>
</tr>
<tr>
<td>5a</td>
<td>Try to link either the initial inner or outer boundary in the new segment with the previous, determined, segment.</td>
</tr>
<tr>
<td>5b</td>
<td>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.</td>
</tr>
<tr>
<td>5c</td>
<td>Check the results of the Markov model for blunders. If blunders are detected, use the original boundaries.</td>
</tr>
<tr>
<td>6</td>
<td>Once a continuous boundary is detected, smooth it with a cubic B-spline.</td>
</tr>
</tbody>
</table>

Table 7.2: These are the individual steps of our boundary detection algorithm presented in a structured outline

### 7.2 ALGORITHM

We will now walk through the algorithm, and we will follow the same OCT image, Fig. 3.4B, through the entire process to illustrate the effects of each step. The following table outlines the individual steps in the algorithm for clarification.
Figure 7.4: This figure illustrates a OCT image after median filtering. The original OCT image is in Fig. (3.4B). The speckle is greatly reduced, and some internal structure is evident within the retina.

Figure 7.5: This figure shows an 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.
7.2.1 Median Filtering (Step 1)

We initially applied a 4x4 median filter twice to each image to suppress the speckle noise. The effect on Fig. (3.4B) appears in Fig. (7.4) and (7.5) for the whole B-scan 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 was far less than that of the speckle.

7.2.2 Column-wise Edge Detection (Step 2)

For our FIR edge detection kernel, we used the second derivative of a one dimensional Gaussian. If \( A \) represents an A-scan, and \( g \) is defined as

\[
g(r) = e^{-\frac{r^2}{2\sigma^2}}
\]  

(7.14)

then

\[
s(r) = g'' * A
\]  

(7.15)

where \( s(r) \) is the output of the edge detection kernel. The edge detector scale is determined by \( \sigma \), the Gaussian standard deviation, and the kernel size was set to be \( 8\sigma \) to limit truncation errors [81]. 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 \( \sigma \) 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're always looking for the same structure, a single-scale analysis is appropriate.

This filter is a one dimensional analog of the Marr-Hildreth operator [82]. As mentioned before, two dimensional edge detectors proved to be problematic due to the frequent dislocations between adjacent A-scans. Other one dimensional, low level edge detectors, such as a one dimensional Sarkar-Boyer [49], could have been used, providing the algorithm was trained with that detector's edges. This edge detector's step response is a zero crossing, whose slope and polarity vary with the size and polarity of the step. A zero crossing response was not required for this problem; an edge detector whose step response was a peak would have served as well. As seen in Fig. (7.5), 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) 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

\[ s_{nzc}(r) = \text{Negative Zero Crossing}(s) \]  

(7.16)
The vector $s_{nzc}(r)$ equals one where $s(r)$ has a negative zero crossing and zero elsewhere.

(2) 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 $s_{nzc} = 1$ was defined as $A \ast g'$. Stronger edges were favored over weaker ones.

(3) 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$^1$ 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

$$s_r(r) = s_{nzc} \frac{g' \ast A}{h_r \ast A}$$  \hspace{1cm} (7.17)

In Eq. (7.17), 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 $1.5\sigma$; 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. (7.6).

$^1$The raw OCT data, as output by the Humphrey 2000, has a positive DC offset and is never zero, even in the dark regions. Hence no problems were incurred by taking this reciprocal.
7.2.3 Initial Edge Selection (Step 3)

The upper and lower retinal boundaries in Fig. (7.6) 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 10 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. (7.7) 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.
7.2.4 Initial Boundary Segmentation (Step 4)

Both initial boundary contours were broken into segments bordered by vertical dislocations larger than 10 pixels in either contour. Jumps larger than 10 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. (7.7), the segment divisions are listed in Table 7.3 below, and are illustrated graphically as the vertical lines in Fig (7.7). The initial locations for both contours are represented by two dots in each A-scan.

7.2.5 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
Figure 7.8: This figure shows the segmentation of the initial approximation to the edges for Fig. (3.4B). The divisions between the segments are the vertical lines, and the boundary location in each A-scan is depicted by a dot. Above the image, each segment is labeled by a number corresponding to the segment numbers in Table 7.3.

<table>
<thead>
<tr>
<th>Segment Number</th>
<th>Segment Length (pixels)</th>
<th>Segment Start (column #)</th>
<th>Segment End (column #)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>88</td>
<td>11</td>
<td>98</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 7.3: This table shows the segmentation of initial boundary contours of the sample image we are using to illustrate the boundary detection system.
within tolerances. Thus, we wish to utilize the instances of small errors and reject the large errors. As a 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 7.3, 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 10 pixels or more.

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 10 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. (7.8). 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.

**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 7.3, we wish to find the boundaries for \( 7 \leq c \leq 10 \), and so using Eq. (7.2) we first predict \( b(10) \) as

\[
\hat{b}(10) = b(11) + \sum_{i=1}^{4} a_i (b(10 + i) - b(10 + i + 1))
\]

(7.18)

where the sum coefficients are from the second row in Table 7.1. Our algorithm then reevaluates the edges responses in the corresponding A-scan for \( c = 10 \). Each edge is given a probability score, \( p \), by

\[
p(r) = g_p(r) s_r(r)
\]

(7.19)

where

\[
g_p(r) = e^{-\frac{(r-k(r))^2}{2\sigma^2}}
\]

(7.20)

and \( s_r \) is from Eq. (7.17). We choose the edge with the largest score and illustrate the outer boundary selection process for a particular column in Fig. (7.9) and (7.10). Fig. (7.9) is a plot of \( s_r \) and \( g_p \); a relatively weak edge is seen to lie close to the predicted location. Fig. (7.10) 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 \( b(10) \) through \( b(14) \) 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, Eq. (7.1) is used. For example, we predict \( b(99) \) from \( b(98) \), \ldots, \( b(94) \) as

104
Figure 7.9: Here is a sample A-scan from Fig. (7.9) showing the Gaussian weighting window centered on the predicted edge location. The boundary we are trying to locate is the outer retinal boundary. The edges are the sharp peaks, where the height of the peak is proportional to the relative edge strength (as defined in Eq. (7.17)).

\[
\hat{b}(99) = b(98) + \sum_{i=1}^{4} a_i (b(99 - i) - b(99 - i - 1))
\]  

(7.21)

where the summation coefficients are now from the first row of values in Table 7.1.

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 \( \sigma_p \) 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. (7.11). If \( \sigma_p \) is too large, then the predicted
Figure 7.10: This figure depicts the A-scan in Fig. (7.9) after the Gaussian window is multiplied by the edge map. The 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.

$b(i)$ 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, the 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.
Figure 7.11: Here is an 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 10, 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.

Normal retinal thickness is never less than 25 pixels (100 $\mu m$) and so our algorithm decides an error is likely if either the final thickness or average thickness is less than 10 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.
Figure 7.12: This figure shows the final determination of the correct boundary locations for Fig. (3.4B). This is the 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.

7.2.6 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. (7.12) 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 (non-overlapping) 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. (7.13), is the final result.
Figure 7.13: Here are the final boundaries after the intermediate results in Fig. (7.12) have been smoothed using the B-spline.

7.3 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 to manually generated ground truth.\(^2\)

We calculated retinal thickness for the \(n^{th}\) A-scan, \(t(n)\), as

\[
t(n) = 4 \frac{\mu m}{pixel} (b_{outer}(n) - b_{inner}(n))
\]

(7.22)

where \(b_{inner}\) and \(b_{outer}\) denote the inner and outer boundaries, respectively and \(4 \frac{\mu m}{pixel}\) 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

\(^2\)Perhaps more properly termed a “target result.”
Figure 7.14: This image shows the 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.

We did not use 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. (7.14) on the far right 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

\[ t_{94} = \frac{1}{94} \sum_{n=3}^{97} t(n) \]  

(7.23)
to be the average retinal thickness over the entire scan (a 94 column ROI), and we defined

\[ t_{33}(j) = \frac{1}{33} \sum_{n=j-16}^{j+16} t(n) \]  

(7.24)

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 corresponds 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.

For the 94 column ROI, the error is

\[ e_{94} = t_{94} - t_{94}^{hc} \]  

(7.25)

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

\[ e_{33} = \max_j \left| t_{33}(j) - t_{33}^{hc}(j) \right| \]  

(7.26)

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.

Intersession variability for OCT retinal thickness measurements has been reported to be on the order of 10 \( \mu m \) [1, 2, 34], which is also the OCT’s theoretical resolution. Thus, we considered thickness measurement errors less than 10 \( \mu m \) (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 \( \pm \) 16 \( \mu m \) [8], and so 10% is 25 \( \mu m \) (6.25 pixels). Therefore we considered errors between 10 and 25 \( \mu m \) to be small. Errors larger than 25 \( \mu m \) were considered to be large.
Table 7.4: This table summarizes the errors produced by our boundary detection system in determining the average retinal thickness of our test images.

### 7.4 Results

*Qualitative:* The boundaries, as determined by our algorithm, generally followed the retinal contours extremely well; the deviations that did occur were most commonly in the outer retinal boundary. Here we have illustrated its performance using as many of the sample Humphrey scans as possible. Some scans required very little or no correction; as in Fig. (7.15). Other scans suffered noticeable dips or rises in the boundaries, causing larger errors. In Fig. (7.16-7.17), the error is still small enough to be classified as *insignificant*. These all come from the Humphrey scans that showed typical errors, and, in particular, Fig. (7.16) features the same scans the Humphrey algorithm had severe errors on. In Fig. (7.18), the error values qualify as "small." Twenty four of the 1450 images, in Fig. (7.19-7.20), contained "large" errors. Fig. (7.20) 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 7.4. The results for the entire set of 1450 images are listed as well as the results from the set of 330 training images.
Figure 7.15: The boundaries in these images had no error.

Figure 7.16: Insignificant error. These are the same two images in Fig. (3.9). In part A, the errors are 0.7 pixels for the 94 column ROI and 1.5 pixels for the 33 column ROI; for part B they are 0.8 pixels for the 94 column ROI and 2.1 pixels for the 33 column ROI.

Figure 7.17: Insignificant error. For part 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. (3.10)
Figure 7.18: Small error. For part 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. (3.8)

Figure 7.19: Large error. In Part A, the errors are 2.3 pixels for the 94 column ROI and 6.3 pixels for the 33 column ROI. Part B features 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.

Figure 7.20: Large error. These two images are among the four worst errors in our testing group. The image in Part A comes from the image set in Fig. (3.10). Part B is 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. The errors are -14.5 pixels for the 94 column ROI and 38.8 pixels for the 33 column ROI.
7.5 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 650 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 edges. These edges could occasionally lead the Markov extension process astray, causing errors like those in Fig. (7.18).

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-dimensional 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. Fig. (7.16–7.17) demonstrate typical 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, Fig. (7.15–7.16), (7.18), and (7.19). 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 [17] and Toth [83]. However, both of these works addressed the false color display of the Humphrey 2000 OCT system, and not the grayscale intensity images used for this study. The bilayer structure is much less visible in the color images. Thus neither work addressed this issue, nor, in fact, did they look at the fovea. Toth's work was near the fovea, however, and her 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 Humphrey algorithm, which, to the best of our knowledge, is the only commercially available system for clinical use. Our algorithm offers fewer errors, and the errors to which it is prone are less severe than those of the Humphrey system. Upon careful inspection of the images presented as acceptable performance of the Humphrey system in Fig. (3.8), one
can note that they exhibit many of the same dips and rises present in the images we presented as having small errors for our algorithm. Moreover, the images presented in Fig. (3.10) as typical Humphrey errors would be classified as large errors for our algorithm. In fairness to the Humphrey system, the two images in Fig. (7.20) that caused the worst errors for our system generated only a typical error for Part A and very little error for Part B.

Finally, any computer vision system will inevitably have errors. The GUI developed to correct the output of our algorithm, 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 of normal, healthy retinas. We expect modifications to be necessary for use with arbitrary retinal images, to address the violations of our retinal model that occur in disease states. For example, some diseases create fluid filled, retinal cysts that are imaged as dark, empty regions within the retina. These spaces may have 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 performance of the algorithm, when extended to the evaluation set, degraded noticeably from the performance over the training set. So long as negligible and small errors are satisfactory, the performance is still very good. However, a larger percentage of the images yield small errors as compared to the
training set. The evaluation set contained a much larger mix of images covering a much wider range of image intensities and hence signal to noise ratios. Also, there were more differences 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.

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 Fig. (7.15–7.20), 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, an OCT system may be capable of detecting and classifying various 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.
CHAPTER 8

A GRAPHICAL USER INTERFACE (GUI) FOR BOUNDARY CORRECTIONS

8.1 Introduction

No computer vision system is 100 percent successful. Images can be found to "break" any system; after all, optical illusions "break" the human vision system. Thus means was needed to make adjustments to the automatically detected boundaries, in anticipation of inevitable errors. The interface must be convenient for the average user to correct individual images and, for our purposes, for us to correct large series of many images. Given these constraints, a graphical user interface (GUI) seemed the most appropriate tool.

8.2 System Description

Our GUI is designed to run under the MATLAB software platform. It is initiated from the command line within MATLAB and can be run in batch mode. The ability to automatically run through a series of images was important, as each of the variability studies generated 150-400 scans to be processed.

Our system opens a window in which the OCT scan is displayed in grayscale, and the automatically determined boundaries are superimposed on the scan. The
boundaries are each stored as a 100 element vector of \( y \) values, where each element represents the boundary location in the corresponding column. Boundary adjustment is initiated by clicking on the image within the GUI. If the location clicked is \((x_o, y_o)\), then the adjustment is made to the boundary location in the columns corresponding to \( x_o \) and its neighbors. The range of the adjustment, i.e. the number of columns to either side of \( x_o \) that are affected, is adjusted by a slide bar on the GUI. The affect is linearly attenuated so that the column at \( x_o \) is given the maximum adjustment and columns furthest from \( x_o \) are given no adjustment. The number of columns affected can vary from 1 to 40, allowing fine adjustments of single columns or coarse adjustments of the entire boundary.

The system can operate in one of three modes. In the first mode, the system assumes that the adjustment is to be made to the boundary whose \( y \) value in column \( x_o \) is closest to \( y_o \). The selected boundary's \( y \) value for column \( x_o \) is then changed to \( y_o \), and linearly scaled adjustments are made to its neighbors as described before. In the second mode, the system assumes that the adjustment is to be made to the outer boundary regardless of what point is selected; the third mode is similar for the inner retinal boundary. The system then uses a b-spline to smooth the final boundary adjustment.

The first mode is usually the best one to use when both boundaries require only small adjustments. The other two are useful in cases where one or both boundaries is severely misplaced and large scale adjustments are necessary. The system stores the original, automatically determined boundaries; thus one can revert to these if severe errors are made during the adjustment process. When the system loads an image data file, the most recently adjusted boundaries are loaded with the image. Finally,
Figure 8.1: A GUI for adjusting retinal boundaries. The controls are the buttons at the bottom.

when adjustments are completed, the system will automatically save the results for the completed image and load the next one specified in its batch job. A screen capture of our GUI is in Figure 8.1
CHAPTER 9

ASSESSING THE IMPACT OF EYE MOTION

9.1 Introduction

Here we propose an algorithm to track the location of the OCT scanning beam relative to the retina. The primary goal of our work here is to develop an algorithm that will report the path of the scanning beam relative to a fixed fundus image of the retina. For this project we will focus on tracking circle scans. As described in the background section, circle scans have a particular sensitivity to ocular motion.

9.2 Theory and Models

9.2.1 The OCT Video General Characteristics

To track the scan path we will rely on the video image of the fundus obtained by the OCT. For the purposes of this research, we will not develop a real-time system to use video as it is obtained. Instead, we are only trying to attain "proof of concept," and so the video output of the OCT fundus video camera is captured by a Hi8 analog recorder and digitized using a Silicon Graphics O2 work station. The video is captured at full resolution using JPEG compression. Despite the compression, the quality is high enough that to a human observer, there is no discernable difference between the uncompressed frames and compressed frames. The video is obtained at the standard
rate of 30 frames per second, and each frame is composed of two fields which are
interlaced. No technical specifications are available for the Humphrey camera. Eye
motion and even the motion of the SLD beam over the retina are fast compared to
the 30 frame per second rate of the camera, and so we will analyze all video as the
individual fields rather than as whole frames. This gives us better time resolution.

The Humphrey video presents us with several challenges. Notably, the quality of
the video as a whole is poor. Our preferred solution to this problem would be to
first change the video camera to a better one. However, this is not feasible. We are
trying to develop a system to improve existing OCT technology. There are currently
hundreds of OCT units sold worldwide, and the new model, which will be released
in 2001 has been reported to have a similar video camera. The existing OCT video
is black and white, but with poor contrast and poor focus. One cause of the poor
contrast is that the camera is sensitive to near-IR; the fundus has particularly low
contrast at these wavelengths. The purpose of this sensitivity is to allow the operator
to view the SLD spot on the retina as the scan is obtained. Two sample OCT video
fields from the same frame are in Figure 9.1.

Relative to the Humphrey OCT, the quality of the video image in Figure 9.1 is
good. Video frames may suffer from occlusion due to subject blinks and defocusing
of the subject's retina. Blinks cause the video frame to become partially or totally
occluded; the occluded portion simply appears black. Normal blinks may last any­
where from one to two frames; one can sometimes see the effects of a blink in one
field but not the other of a given frame. Defocusing often happens during the one or
two frames bracketing a shift in a subject's eye position. In Figure 9.2 is a sample
series of video fields that illustrate this defocusing.
Figure 9.1: Here are sample video fields as captured from two different subjects using the Humphrey video camera. Note the low contrast and poor focus.

Figure 9.2: This series of video fields, obtained in sequence, demonstrate how the image can suddenly blur. The middle image in particular shows almost no features.
The optic nerve head is the disk just above and to the left of the center in Figure 9.1A, and several blood vessels can be seen as the dark ribbons emerging from it. The SLD spots are the bright white regions that are just to the left of the nerve head and far to its right. In Figure 9.1, we see the graininess and lack of contrast that plagues the normal OCT video. Also, the video saturation caused by the SLD spot can be seen.

The appearance of the retina is not constant from subject to subject. Pigmentation, the relative size of the nerve head, and the blood vessel pattern may differ markedly between subjects. Retinal pigmentation tends to be associated with skin pigmentation, with more pigmented peoples having more pigmented retinas. The size of the nerve head image in the fundus video also varies among people. This is, in part, due to actual size differences in nerve heads; however, part is also due to differences in ocular length and refraction. Finally, the blood vessel pattern is a random feature that is unique to each individual, within a few restraints. In Figure 9.3 one can see a sample video field from each of the six subjects used in this study.

9.2.2 The SLD Spot

For nerve head scans, one can assume that the SLD spot travels in a circle about the nerve head. In Figure 9.4 we have the final output picture obtained by the Humphrey, showing the circular path drawn around the optic nerve head. Note that Figure 9.4 is obtained with the red free light, and thus has much higher contrast and many more blood vessels.

The identification of the SLD spot poses several challenges. First of all, the SLD illuminates 66 retinal A-scan locations per second, while the video camera obtains 60
Figure 9.3: Here is one sample video frame from each of the six training subjects used in this study.

Figure 9.4: Here is a scan image showing the circular path a scan may take around the optic nerve head; the path is illustrated as the bright circle. This view was created by the Humphrey system *after* the scan was acquired.
fields per second. Thus the two are not synchronized, and so a given field may capture
the locations of one or two A-scans. This fact, when coupled with the seeming slow
shutter in the OCT camera makes the appearance of the scanning spot vary from
field to field. Another contributing factor to the variation of SLD appearance is
the non-Lambertian nature of the retinal surface. The retinal surface can return
specular reflections from the scanning location as well as elsewhere on the retina.
This causes additional variations in the appearance of the scanning spot as well as
additional bright spots on the retina that can be mistaken for the SLD. Finally,
the SLD beam serves to illuminate a landmark spot on the retina in addition to
obtaining the A-scans with the scanning spot. To a scan operator watching the live
fundus video, this landmark spot appears as a fixed bright spot on the retina as the
scan is obtained; however, it is generated by the OCT periodically switching the SLD
from the scanning spot to the landmark spot. One can see this switching when the
individual fields are viewed; in a given field, one may see just the landmark spot,
just the scanning spot, or both spots. As for the cases in which both spots are
visible, it is not clear whether the switch in SLD location occurs during the middle
of a video field, so that both locations appear illuminated in the recorded image, or
whether the SLD beam switches back and forth very rapidly. The existence of fields
in which only the landmark is visible suggests the former, however. Thus with the
landmark light and the possibility of glare, any given video field captured during scan
acquisition may have several bright spots which must be examined and identified.
Between scan acquisitions, the SLD stops its scanning motion, and either disappears
entirely or illuminates a circle showing the intended scan path. Thus we must be able
to contend with the disappearance and reappearance of the scanning spot as well.
The detection and identification of the scanning SLD spot as well as the landmark SLD spot can be made simpler if we model their behaviors. The scanning SLD moves in a circle; in particular, its (row, column) position in the image matrix \((r, c)\) can be described by the locus of points satisfying

\[ R^2 = (r - r_c)^2 + (c - c_c)^2 \]  

(9.1)

where \( R \) is the radius of the circle and \((r_c, c_c)\) is the center of the circle. The OCT performs always performs a circle scan in the same period of time, approximately 1.5 seconds. Thus the scanning path can be expressed parametrically as

\[
\begin{align*}
    r &= R \sin(\theta_o t) + r_c \\
    c &= R \cos(\theta_o t) + c_c
\end{align*}
\]  

(9.2)

where \( \theta_o \) is a fixed angle which the SLD beam travels between each field. The parameter \( \theta_o \) can be estimated a priori based on knowledge of how the Humphrey OCT
system works. However, the parameters $R$, $r_c$, and $c_c$ cannot be known in advance, as they may differ from one scan to the next. Moreover, the operator may adjust these in the middle of a scan.

If a given set of SLD positions can be hypothesized to be the scanning SLD beam, however, then an estimate of these parameters can be made. In particular, for three points, $(r_1, c_1)$, $(r_2, c_2)$, and $(r_3, c_3)$, the unique circle that passes through all three can be calculated as follows. First note that each satisfies Eq. 9.1. In particular, one can combine the equations for $(r_1, c_1)$ and $(r_2, c_2)$ to yield

$$r_1^2 + c_1^2 - r_2^2 - c_2^2 = 2(r_1 - r_2)r_c + 2(c_1 - c_2)c_c \quad (9.3)$$

Similarly, one combines the equations for $(r_1, c_1)$ and $(r_3, c_3)$ to yield

$$r_1^2 + c_1^2 - r_3^2 - c_3^2 = 2(r_1 - r_3)r_c + 2(c_1 - c_3)c_c \quad (9.4)$$

and thus solves for $(r_c, c_c)$. Once these are known, $R$ can be found.

Given a set of points $(r_i, c_i)$ hypothesized to be the SLD scanning point, one method of finding the best circle estimate for the scanning path is to use a least squares approach. This can be problematic, however, because outliers can have undue influence on the result. A more robust technique is similar to the RANSAC technique [84] used by Can et al [58]. In our implementation, one randomly chooses triplets of points from the set $(r_i, c_i)$ and finds their individual circles. The final circle is the median of the results for all triplets.

Localizing the landmark SLD spot is somewhat simpler in that it generally does not move in the course of a scanning session. While it can be relocated after scanning has started, doing so requires the re-initialization of the OCT scanning software; thus
any such relocation could be considered a new scanning session. The landmark is characterized both by its spatial invariance and its particular temporal pattern of instantiation. The landmark is not present in every video field, and a careful study of video fields shows that its pattern of presence and absence can be modelled as a square wave which shows the landmark as being on for two fields and then off for three. This integer ratio of durations, expressed in video frames, is not precisely correct because the SLD beam is not precisely synchronized with the video camera. Thus one will occasionally find the duration of the off phase to be two fields instead of three. However, we ignored this discrepancy, because the landmark spot was still very close to this 2:3 duty cycle, and no other spot in the field was.

9.2.3 The blood vessels

The retinal blood vessel patterns vary from subject to subject as can be seen in Figure 9.3. However, some general rules hold. First of all, the vessels are arranged into superior and inferior arcades that extend above and below the nerve head, respectively, and then branch out to either side. These branches generally encircle but don't encroach on the fovea, where the central visual field is perceived; the fovea is located to the temporal (i.e. away from the midline of the body) side of each nerve head. There are actually two vascular trees, an arterial and a venous one, but the image quality is not sufficient to reliably distinguish or even detect both of these, and so our approach will be to simply look for vessels.

The blood vessels are locally dark ribbons, though they are not necessarily the darkest regions in the image. Moreover, there may be specular reflections from the vessel. Though far less bright than glare from the SLD, these reflections can give
These images show the appearance of the blood vessels in the video images. The left image shows the entire video field, with a smaller region selected in the box. The middle image shows this box close up, while the right image shows a single, representative column from the middle image.

blood vessels a hollow appearance. The blood vessels can be considered to be smooth contours that may branch; however, a sufficient number of branch points cannot be reliably detected for feature based tracking and so they will be ignored. As the vessels move further from the nerve head and branch out, they become thinner. The OCT camera can only reliably show the largest vessels within two disc diameters of the nerve head; the vessels soon dissolve into the noise and local retinal pigment variations. Our detector will thus be tuned for large vessel sizes.

To help remove the noise, we condition our images with a minimum filter, as described by Can et al [58]. The minimum filter works like a median filter, except in each local neighborhood it chooses the minimum pixel intensity. The effect is to dilate the vessel profiles somewhat and give them more uniform, dark interiors. In Figure 9.6 we have an enlarged view of a blood vessel as well as a representative cross section along one image row.

The large retinal blood vessels that we will be detecting exhibit some organization to their geometric arrangement relative to the nerve head; they are not arranged in
Figure 9.7: The vessel distribution in these images is a pattern for our idealized, radial distribution of blood vessels. Note how many large vessels appear to come out from the nerve head center. Note too that many vessels don't follow this pattern.

a haphazard way. Each subject's retina is different, but in a general sense, the large blood vessels extend radially out from the retina. This is necessary anatomically because the vessels emerge onto the retina from within the optic disk and then branch out to nourish the rest of the retina. Our paradigm for vessel distribution will be that in the figure below.

Of course, different subjects resemble this paradigm to different degrees, as can be seen in Figure 9.3. In addition to the between-subject differences in vessel distribution, the differences in magnification between subjects causes different amounts of the tree to be visible in different subjects.

9.2.4 The Optic Nerve Head and the Retina

The specific appearance of the optic nerve head differs from subject to subject, as evidenced by the images in Figure 9.3. The same can also be said of the retinal regions that do not contain the nerve head. However, the general appearance of the
nerve head is a locally bright ellipse. The normal nerve head is known to be slightly elongated in the vertical (superior-inferior) direction. However, the effect is small and is apparently countered by the aspect ratio of the pixels in the full frame image so that the pixel dimensions of the nerve head appear to be the same in both the vertical and horizontal directions. When the frames are decomposed into individual fields, then the number of pixels becomes twice as large in the horizontal direction. To compensate for this we also reduce the column resolution of each field by two. Though this reduces the information available in the image, it does not seem to adversely affect performance and serves to reduce the number of pixels that must be processed by our tracker. Similarly, the retinal regions that do not feature the nerve head are generally of far less contrast and devoid of coherent features other than blood vessels. Because we are interested in locating the optic nerve head, we will call these locations "empty retina." One caveat to this description is the SLD spot; its presence can give a retinal location a very bright, roughly circular spot. However, there are means that will be discussed to reduce its interference in our detection scheme.

Because of the general similarity of nerve head and retinal appearance among subjects, an eigenimage based approach appears to be a suitable recognition strategy. In particular, local \( m \times m \) subimages can be selected that are just large enough to contain the expected range of nerve head sizes. All possible subimages fill the image space \( R^M \), where \( M = m^2 \). Our problem domain is restricted to fundus photographs, and so we can consider only two classes of subimages, those that are centered on the optic nerve head and those that are centered on "empty retina." Within \( R^M \), then, subimages centered about the optic nerve head can be said to occupy a region \( \Phi_N \) while those centered on empty retinal regions occupy the region
As with the general eigenimage approach, these regions can be approximated as separate subspaces of $R^M$ and a principal components analysis can be performed to find the basis vectors for each subspace that contain the most energy. Thus for the two regions $\Phi_R$ and $\Phi_N$ we construct the subspace approximations $\hat{\Phi}_R$ and $\hat{\Phi}_N$ that are the span of the highest energy eigenvectors. A selected test subimage can then be projected into both $\hat{\Phi}_R$ and $\hat{\Phi}_N$; the relative energy of the two projections can then be used to determine the likelihood of the template containing the optic nerve head.

Two points should be mentioned. First, there is the anatomically obvious fact that there is only one optic nerve head per eye (normally!). The non-trivial implication, however, is that one ultimately has only to choose the single most likely optic nerve location for each frame. This differs from many other eigenimage based recognition strategies where there may be multiple targets to locate. The second point concerns the use of the projections to choose templates centered on the nerve head. To help ensure that only those templates gave a strong response, the “empty retina” training set included templates which contained a portion (up to half) of the nerve head.

### 9.2.5 Ocular and SLD Motion

We are trying to track both the SLD beam and the optic nerve head as both features move during the course of B-scan acquisition. The SLD beam is straightforward in that it follows a highly regular path. Thus once the parameters of the path are known, the SLD beam location can be predicted from field to field. The nerve head, however, does not move in a predictable manner. There is usually a small amount of “jitter” from field to field, as the nerve head makes shifts of up to 0.25 nerve head diameters ($nhd$). However, one can also have large saccades where the nerve head
moves suddenly and unpredictably in any direction. These saccades can be of large magnitude relative to the small viewing angle on the video monitor; the nerve head can easily change its position by half the screen size or more. The motion is fast relative to the video field capture rate; motion across half the screen can occur over only one or two fields. Because of the blurring that is often associated with motion, the effect is that the nerve head is in one location for a period of time, "disappears" for one or two fields when the image is blurred, and "reappears" in a completely new location in the next field. Thus traditional trackers which rely on inertial motion are of limited use. One can predict the nerve head to be close to previous locations; however, the possibility exists that it will suddenly be in a completely new location.

9.2.6 Dual Class Eigenimage Recognition

Eigenimage approaches have been used in the literature to recognize objects in images. Typically, the class objects to be recognized form a set, $C$, which may be the union of several subsets so that $C = \bigcup_i C_i$. For example, the set of facial images to be recognized, $C$, may be composed of many subsets, $C_i$, where each $C_i$ is the collection of images from a different person. For recognition purposes, then, one can construct the eigenspace $\Phi_C$ for the image class $C$; an image, $\tau$, is evaluated then by projecting it into $\Phi_C$ and measuring the energy in the projection. If the energy exceeds some threshold, this is interpreted as implying that $\tau \in C$. One can also look at the projection of $\tau$ into $\Phi_C$ and compare it to the projections of elements of the various $C_i$; if the projection of $\tau$ lies close to that of some $C_j$, then we can further conclude that $\tau \in C_j$. 

137
In typical detection/recognition problems, an image \( I \) is much larger than the size of the object being sought. The approach used is to look at subregions \( \tau \) that are just large enough to contain the images of the object and let \( C = \{ \tau \} \). The eigenspace \( \Phi_C \) is then created from the images in \( C \) in the usual way. To detect the class \( C \) within a given image \( I \), then, one scans all local neighborhoods, \( \tau^* \), of size \( \tau \) within \( I \). Each \( \tau^* \) is projected into \( \Phi_C \) to decide whether or not \( \tau^* \in \Phi_C \). In many detection problems, the null hypothesis that \( \tau^* \notin C \forall \tau^* \) is a valid conclusion. Thus if no \( \tau^* \) projects strongly enough, the conclusion is \( C \) does not occur in \( I \).

In this problem domain, however, the null hypothesis is unlikely so long as the subject is cooperating and has their eyes open. It will be assumed during the one or two seconds of the adaptive phase that the subject is able to keep their eyes sufficiently fixated so that the nerve head is visible in all video fields. Thus for each video field we find the most likely nerve head location and assume that the nerve head is actually there.

Another characteristic of this problem is that our background is fairly uniform. Though there are retinal blood vessels and retinal pigment variations, these are of low contrast. In many other detection domains, the background can have many random, unpredictable variations with large contrast; in these cases, one can only consider the class of known objects, \( C \), that is being detected. In this case, however, it is possible to consider two classes, \( C_R \) and \( C_N \) that are the class of \( \tau \) containing "empty retina," i.e. retinal locations without the nerve head, and the class of \( \tau \) centered on the optic nerve head. Thus the corresponding eigenspaces, \( \Phi_R \) and \( \Phi_N \) can be constructed.

To determine whether the nerve head is located at a given location, \( (\tau_i, c_i) \), then, a subregion \( \tau^* \) is created centered on \( (\tau_i, c_i) \). This \( \tau^* \) is projected into \( \Phi_R \) and \( \Phi_N \),
generating projection magnitudes of $\pi_R$ and $\pi_N$, respectively. We can consider $\pi_R$ and $\pi_N$ to be fuzzy membership functions of $\tau^*$ into $\Phi_R$ and $\Phi_N$; thus $\tau^*$ is considered to contain the nerve head if $\pi_N > \pi_R$. In particular, we evaluate the ratio

$$\eta = \frac{\pi_R}{\pi_N}$$

as a measure of the difference in membership. Thus, to locate the optic nerve head, we evaluated candidate retinal locations based on $\eta$; higher values were interpreted as more surety in identification. Comparing the ratio $\eta$ proved to be more reliable than the individual membership scores $\pi_R$ and $\pi_N$.

To create the eigenspaces $\Phi_N$ and $\Phi_R$, 200 video fields from six different subjects were manually marked to indicate the center of the optic nerve head. The nerve heads frequently had some vascular structures that allowed the same point to be reliably marked, with an error of only a few pixels. Note that this point wasn’t necessarily in the center of the nerve head. In each field, then, we formed a training subimage $\tau_N$ and from all the fields, then, we formed $C_N = \{\tau_N\}$. In each field we also made training subimages $\tau_R$ of empty retina, and combined them to form $C_R$. The $\tau_N$ were large enough to contain a ring of retina around the nerve heads to help make the nerve heads more recognizable as a feature on the surrounding retina. The $\tau_R$ were selected by randomly selecting six points within each field that were sufficiently far from the nerve head center. In order to improve the localization of the nerve head, the $\tau_R$ were allowed to contain up to half the nerve head. The selection was done in the images after the removal of the SLD spot, which will be discussed in later sections. Samples of these training subimages from each subject are shown in Figures 9.8 and 9.9. In
Figure 9.9, one can see that some of the training subimages for $\Phi_R$ contain portions of the nerve head.

To improve system speed, the resolution of the images and hence the $\tau$ was reduced by half in both the row and column direction. Since each field as processed by the other steps had half as many rows and columns as the original video frame, this reduction represents a factor of four over the original video frame. This lower resolution preserved sufficient detail for the nerve heads to be recognizable. The size of our training subimages in this reduced resolution was 43 x 43 pixels; typical nerve head diameters ranged from 30 to 40 pixels.

Once our training subimages were collected, the corresponding eigenspaces $\Phi_R$ and $\Phi_N$ were created. The 43 x 43 training subimages correspond to an eigenspace with 1849 dimensions. We had approximately 1200 nerve head images and 6000 empty retina images (the positioning of the nerve head and subject blinks did not always allow the collection of six empty retina images from a video field). One reason so many images of the nerve head were collected is because eigenimage techniques are susceptible to changes in illumination. In the retinal video, however, the movement of the SLD scanning spot causes large changes in the direction of illumination. The large number of images allows samples to be used from many different SLD positions, and these differences in illumination can be seen in Figure 9.8. A similar situation exists with the retinal images. By using a random selection technique, we acquired images with the SLD spot in different locations in the retina and within the template.

Unlike many eigenimage implementations, the number of training images is on the order of the image space dimensionality. However, we must choose a smaller dimensionality so that our system can generalize and also be fast. Thus we chose
Figure 9.8: Here are some sample training subimages used to train the nerve head eigenspace. Shown here are two sample training images from each subject.
Figure 9.9: Here are some sample training images used to train the retinal eigenspace. Again, we have two sample images from each subject.
a smaller set to approximate $\Phi_R$ and $\Phi_N$. We did not choose on the basis of the
eigenvalues $\{\lambda_R^i\}$ and $\{\lambda_N^i\}$ for the corresponding sets of eigenvectors $\{\phi_R^i\}$ and $\{\phi_N^i\}$. The $\phi_N^i$ with the largest $\lambda_N^i$ are best at representing the $\tau_N$, but not necessarily in discriminating between $\tau_N$ and $\tau_R$. Instead, we systematically searched for the most
discriminating set of eigenvectors in $\{\phi_R^i\}$ and $\{\phi_N^i\}$.

A given combination, denoted as the $j^{th}$ combination, can be represented as
$\{\phi_R^{m,j}\}$ and $\{\phi_N^{m,j}\}$, where $\{n_i^{m,j}\}$ and $\{r_i^{m,j}\}$ denote which specific nerve head and retinal eigenvectors are included. Here the $i$ indexes the eigenvectors for the $j^{th}$
combination, and the $m$ simply labels this combination as having $m$ eigenvectors in
each eigenspace. Note that the $j^{th}$ combination encompasses both the retinal and
nerve head vectors. Also, in general, $\{n_i^{m,j}\} \neq \{r_i^{m,j}\}$, though they both have the
same number of elements. This wasn't a formal constraint implied by our model.
However, if either $\Phi_R$ or $\Phi_N$ had more eigenvectors than the other, images tended to
project better into that space. Hence the combination of $\Phi_R$ and $\Phi_N$ would be poor
at distinguishing between empty retina and nerve head.

The resulting eigenspaces are denoted $\{\hat{\Phi}_R^{m,j}\}$ and $\{\hat{\Phi}_N^{m,j}\}$. A given combination
was evaluated by projecting all the $\{\tau_R\}$ and $\{\tau_N\}$ into $\{\hat{\Phi}_R^{m,j}\}$ and $\{\hat{\Phi}_N^{m,j}\}$ and cal-
culating the resulting $\eta$. A retinal template $\tau_R$ was correctly identified if $\eta < 1$ and
likewise a nerve head template was correctly identified if $\eta > 1$. Each combination of
eigenvectors was given the score $c_{m,j}$ which was its error rate over all the $\tau_R$ and $\tau_N$.

We did not know the optimal dimensionality $m$ for $\{\hat{\Phi}_R^{m,j}\}$ and $\{\hat{\Phi}_N^{m,j}\}$, and a thor-
ough search of all possible combinations of eigenvectors was prohibitively expensive
for even small dimensions due to the sheer combinatorics. Thus, for a given $m$, we
cannot explore all possible $\{\phi_R^{m,j}\}$ and $\{\phi_N^{m,j}\}$. Our solution was to try a modified
search. For \( m = 1 \), i.e. one eigenvector per space, we found the optimal pair \( \phi_R^1 \) and \( \phi_N^1 \) that minimized \( c_1 \). This optimal pair is then denoted as \( \{ \phi_R \}^1 \) and \( \{ \phi_N \}^1 \). For \( m = 2 \) we then added a second pair of eigenvectors, and thus found the optimum \( r_2 \) and \( n_2 \) so that \( \{ \{ \phi_R \}^1, \phi_R^2 \} \) and \( \{ \{ \phi_N \}^1, \phi_N^2 \} \) have the minimum value of \( c_2 \). This process continues iteratively, so that at the \( k^{th} \) step we find the optimum \( r_k \) and \( n_k \) so that \( \{ \{ \phi_R \}^{k-1}, \phi_R^n \} \) and \( \{ \{ \phi_N \}^{k-1}, \phi_N^n \} \) have the minimum value of \( c_k \).

The values of \( c_k \) did not decrease monotonically with \( k \) for all \( k \). Instead, a minimum was reached at \( k = 15 \), after which the error increased. A plot of this is in Figure 9.10.

The optimal \( n_i \) were \{1, 2, 6, 13, 25, 9, 5, 3, 15, 19, 29, 24, 17, 20, 11, 30, 27, 28, 10, 22\} and the optimal \( r_i \) were \{1, 2, 7, 6, 12, 5, 9, 3, 26, 13, 22, 11, 19, 23, 18, 25, 28, 30, 8, 29\}. These were the order in which they were chosen. Furthermore, the choices were limited to the eigenvectors corresponding to the largest 30 eigenvalues in each space, respectively, to speed our search. Note that order of the choices did not strictly follow...
Figure 9.11: Here are the eigenvectors representing $\hat{\Phi}_N$

the energy in each eigenvalue. The resulting eigenvectors corresponding to $n_i$ and $r_i$, in the order they were chosen, are in Figures 9.11 and 9.12. The eigenvectors for $\Phi_N$, as expected, look somewhat like generic nerve heads. The eigenvectors for $\Phi_R$ are more interesting, in that they look like a very general set of gradients and two dimensional sinusoids. This is probably because the retinal templates do not have a consistent structure, and so these eigenimages represent the most efficient subspace for general images.

One can see the discriminating ability of the dual space projection technique in the following sample video fields in Figures 9.13 - 9.18. For each field, subregions are chosen centered around every image pixel that is sufficiently far from the edge. These
Figure 9.12: Here are the eigenvectors representing $\Phi_R$
subregions are projected into $\hat{\Phi}_N$ first, and then into $\hat{\Phi}_R$. In these images, darker regions imply a larger magnitude of the projection. Finally, $\eta$ is calculated for every possible image pixel by calculating the ratio of the two projections, and the values of $\eta$ are also displayed as a surface plot for additional clarity.

9.3 System Description

9.3.1 General Overview

The tracking algorithm has two separate parts, a system for tracking the SLD spot and a system for tracking the optic nerve head. Both systems operate in three modes. First there is an adaptive phase, where the systems gain information and make tentative identifications. As more information is gained, both systems then characterize the features of their targets so as to make more efficient searches in later frames. Once information is known, the systems enter the operational phase. Finally, both systems reassess their identifications when their targets are not found in their expected locations. If a sufficiently large discrepancy is found, then the systems assume that some of their assumptions are wrong and they reinitiate adaptation.

The SLD tracking system works by finding the brightest points in each image. A clustering algorithm is used to associate these with candidates for the laser spot in each field. After analyzing the positions of the candidates over several fields, decisions are made as to which correspond to the SLD scanning beam and which correspond to the SLD landmark and these are then tracked over each subsequent field.

The nerve head tracking system is more complicated. The tracking is ultimately performed by detecting the retinal blood vessels and possibly the nerve head boundaries, and following them from frame to frame. The blood vessels move rigidly with
Figure 9.13: This image shows how the dual eigenspace approach is able to localize the nerve head in a video image. On this and the next few pages, we show two sample fields for each subject. We start with the video field in the upper left, followed by $\pi_N$ and $\pi_R$. The bottom row show $\eta$, in intensity form and mesh form for emphasis. The video images are shown with a reverse color map, so that darker implies larger values of $\pi_N$, $\pi_R$, and $\eta$. 
Figure 9.14: Here are more samples.
Figure 9.15: Here are more samples.
Figure 9.16: Here are more samples.
Figure 9.17: Here are more samples.
Figure 9.18: Here are more samples.
the nerve head, and so tracking them is equivalent to tracking the nerve head. The blood vessels are the most reliable features that can be detected in each frame, though only a small portion of the vascular tree is likely to be visible on any given field. The initial set of video fields in the *adaptive* phase are used to build a model of the retinal vascular tree. Once a satisfactory model is built, then the detected vasculature in each subsequent field is compared to this model to determine the relative position of the nerve head for the *operational* phase. To build a coherent model combining the features found in different video fields, we must be able to find the correct correspondences and geometric relationships between them. The common reference point we will use in each video field during the model building phase is the detected nerve head position.

The nerve head detection in these video images is a very difficult problem, for the reasons stated before. We have not found any single technique that is sufficiently reliable to be used by itself. Instead, we have found a combination of three techniques, which, when used together, successfully detect the nerve head the majority of the time. Each technique is liable to preferentially select a false target over the actual nerve head. However, the set of false targets for each technique is different than the sets of false targets of the others, though they are not disjoint. Even with the combination, errors are still made. However, as will be discussed, the errors occur less frequently and less coherently than the correct detections, allowing a map of the retinal vasculature to be created.

The combination works in three stages, where each stage limits the search region of the next. The stages are arranged strategically to best complement each other's strengths and weaknesses. The first stage filters the fields with a fast 1-D filter that
selects the vessels and the nerve head boundaries. A chain coder finds the retinal blood vessels and the nerve head boundaries from the filter results, and a Hough transform is then used to find circular candidates for the nerve head. Only the regions indicated by these circles are sent to the second stage, which uses the eigenimage approach to evaluate them. In particular, it projects them into $\hat{\Phi}_R$ and $\hat{\Phi}_N$ and compares the relative projection magnitudes. Only the most promising circles are then evaluated using the third stage. This stage analyzes the geometric relationship between the candidate circles and the detected blood vessels and nerve head boundaries. Once the final nerve head hypothesis is reached, the relative positions of the detected blood vessels are noted and they are added to the growing model.

Once the model is sufficiently stable and descriptive, the adaptive phase is ended and subsequent tracking and detections are done using only this model and the detected blood vessels found by the chain coder. These vessels are compared to the model and the nerve head position is inferred.

9.3.2 Initial Image Acquisition and Conditioning

The initial image acquisition and conditioning is the same regardless of whether the algorithm is still adapting or actively tracking. As stated before, the OCT video output was saved on Hi8 analog film and then digitized using a Silicon Graphics O2 workstation. The video frames are split into their individual fields and each field is then analyzed to determine whether it has a valid retinal image or whether it has been obscured by a blink. The test for field validity is a simple threshold test; if the mean pixel intensity over the image is below the threshold $i_o$ then the image is assumed to be obscured. This threshold was found empirically; typical mean pixel
intensities range from 70 to 120 units/pixel, and so \( i_0 \) was set to 30 units/pixel. If a field is determined to be invalid, it is skipped. Otherwise, the next step is a minimum filter. This nonlinear filter is analogous to a median filter. For each image pixel, the \( n \times n \) neighborhood around that pixel is analyzed, and the pixel value is set to the minimum intensity in that \( n \times n \) neighborhood. We used a \( 3 \times 3 \) minimum filter.

As can be seen in Figure 9.19, the minimum filter serves both to remove noise as well as to "fill out" and expand the blood vessel images. This allows them to have better boundaries for detection by the 1-D filters used for the nerve head detection.

### 9.3.3 SLD Spot Detection and Tracking

**Locating all spots within each field**

We will first describe the SLD tracking subsystem, as it is somewhat simpler than the nerve head tracking subsystem. The first step, after the initial image conditioning, is always to locate the SLD spot and its resulting glare in the field. During adaptation, when the algorithm hasn't learned to distinguish between these objects, information is simply gathered on all bright points within the video field.

The SLD spot, when it is present, is remarkably consistent in saturating the CCD camera within the OCT. This is true for the SLD scanning spot, the SLD landmark spot, and glare resulting from the SLD. The pixels within the spots are always the brightest pixels in the image, and are usually two to three times the intensity of the surrounding retinal pixels. One simple approach which works remarkably well is to choose the brightest \( N \) pixels within a field; here \( N \) is set empirically to be 100. This approach was chosen rather than an intensity threshold because the SLD pixels do not always exhibit such an extreme intensity difference between themselves and the
Figure 9.19: Here are some sample fields showing the effect of the minimum filter (pre-filter is on left, post-filter is on right)
neighboring retinal pixels. Once pixels were selected, a simple clustering algorithm groups them into clusters.

For this particular problem, a maximum cluster size \( r \) was chosen based on the normal appearance of the SLD spot. The size can vary markedly, depending on the degree of focus and the amount of glare present. A value of \( r \) was chosen in favor of reducing the incidence of multiple responses from an SLD spot while retaining sufficient resolution to distinguish between spots. This does require compromise because the SLD scanning spot can move very close to or even over glare or the SLD landmark spot. The final value chosen was 1.5 times the typical SLD spot diameter. The clustering algorithm works in two iterations. Say the set of selected pixels is \( \{x_i\} \), where \( x_i \) represents a point in the field, then for the first iteration, the first cluster, \( s_1 \) is defined as the median of \( \{x_i\} \). If \( \{x_i\} \) has an even number of points, additional code ensures that \( s_1 \) equals one of the \( \{x_i\} \). If all points in \( \{x_i\} \) lie within \( r \) pixels of \( s_1 \), then the algorithm stops. Otherwise, the most distant pixel is chosen as \( s_2 \). This process continues until all of the \( \{x_i\} \) lie within \( r \) pixels of some \( s_j \).

A problem with this first iteration is that we are not guaranteed that the central pixel of the SLD spot will be among the initial \( \{x_i\} \), and even if it is, it may not be picked as the center of a cluster. Thus, for the second round, the clustering algorithm looks in the neighborhood of radius \( r \) about each \( s_i \) and chooses the average, intensity weighted pixel location. If image pixel intensity is imagined to be a mass density function, this is akin to choosing the center of mass and will be referred to as such. Choosing the center of mass is superior to choosing the brightest pixel because the SLD spots may have a large number of saturated pixels. Also, the pixels at the center of the SLD spot are often so much brighter than all other image pixels that
Figure 9.20: Here are some sample fields showing the clusters that were organized into bright spots by the SLD finder subsystem.
the calculated center of mass is usually close to the center of the SLD spot regardless of the initial location of $s_j$.

Labeling the spots and adaptation

Once the bright spots have been discovered in a given field, the next step is to label them. From the data in only one video field, it is difficult to distinguish glare from an SLD spot, and impossible to distinguish between the SLD scanning spot from the SLD landmark spot. However, the SLD scanning spot and SLD landmark spot have regular, predictable behavior over time. Thus the strategy of the adaptive phase is to label all bright spots in each video field, noting correspondences between spots in sequential fields to give consistent labels. By observing the temporal behavior of the labeled spots, their identities can be determined. Labeling is done at all stages of our algorithm. However, during the initial adaptive phase, all detected spots are labeled without any information as to which are the SLD spots. Once the identifications are made, however, only the progress of the SLD spots is recorded over time.

In the first field during the adaptive phase, each bright spot are given its own label. In all subsequent fields, starting with the second one, the spot locations are first compared to the labeled spot locations in previous fields. If the spot location in the current field is sufficiently close to an instance of a previously labeled spot, it is considered another instance of that spot and assigned the same label. A separation of 25 pixels or less was considered sufficiently close, which takes into account the expected distance the SLD scanning spot can travel between fields, as well as the typical error range in identifying the center of a bright spot. Note that a labeled point does not have to occur over subsequent fields; the algorithm can look back over all the labeled fields to determine if a bright spot has occurred before at a given
location. The SLD landmark spot in particular does not occur in every video field, and glare also comes and goes depending on the position of the SLD beam. In Figure 9.21, the labels given to the spots in individual fields are shown. The spot labeled 1 in both fields is actually the SLD scanning spot, the one labeled 2 is actually glare, and the one labeled 3 is the SLD landmark spot.

**Identifying the labeled points**

The data gathering occurs over 12 fields, a number that was set empirically. Twelve fields is enough for the behaviors of the SLD spots to be noticeably different than those of the glare. In particular, this is sufficient time for the SLD scanning beam to travel approximately 40° and to have allowed the SLD landmark spot to have flickered on and off for at least two cycles.

We first identify the SLD scanning spot. The model used is that it is the only spot with consistent motion between fields. The SLD landmark spot remains in the same location by design, and glare is also considered to occur at fixed locations. When
glare appears in different locations, it is considered to be instances of different glare spots. All spots will have some motion because their shapes generally change from frame to frame and so their centers of mass may move. Glare spots in particular may have irregular shapes so that their centers of mass are not well defined and vary between frames. Nonetheless, these motions will be random and the net motion from field one to field eight should be small. The SLD scanning spot, on the other hand, moves in a consistent way and the expected displacement over twelve fields can be estimated. There is some variation because the displacement depends on the radius of the scanning circle. However, a reasonable minimum can be calculated, and so the SLD scanning spot is identified as the labeled spot featuring the largest displacement that is larger than this reasonable minimum. To guard against a chance misidentification of glare, the proposed SLD scanning spot is further required to have regularity in its motion so that its displacement during the first four fields is approximately equal to its motion in the last four.

The SLD landmark spot is then identified based on its temporal behavior. A matrix showing the temporal behavior is in Figure 9.22. In the figure, the horizontal axis denotes time as counted by video fields, while the vertical axis keeps track of the individual spots identified by the labeling. Each row thus represents an existence matrix for that spot; a one (white) in a particular column indicates that the spot was detected in the corresponding video frame.

Labeled spots come and go from field to field, either because they do not consistently exist in subsequent fields or because the spot detection algorithm missed them. Glare in particular tends to have a random temporal existence depending highly on the precise location of the SLD spots. The SLD scanning spot, in contrast,
Figure 9.22: Here is a matrix indicating the video fields in which each labeled spot existed; white means existence, black means not existing. The bottom row is the landmark spot; note the regular pattern. The top row is the scanning spot; note that it is there for most fields (the middle row is glare).

is usually present, though it does disappear occasionally for a single field. Its periodic disappearances are tied to the illumination of the SLD landmark spot, which has a strikingly regular temporal behavior. In particular, it is predictably on for two fields and then off for three. This cycle is not perfectly tied to the video camera; occasionally one sees the SLD landmark being off for only two fields or on for just one. Nonetheless, it is the only labeled spot expected to exhibit this behavior. Thus the average on and off times are calculated for all labeled spots (other than the SLD scanning spot). Each labeled spot has its average on and off times arranged into a vector; for the $i^{th}$ label this is

$$t_i = (\bar{t}_{on}, \bar{t}_{off})$$

(9.6)

and the spot closest to $(2,3)$ is considered to be the SLD landmark spot. This distance is measured as a simple geometric distance in two-dimensional space. In

163
Figure 9.22, we see that spot number 3 is closest to being on for two and off for 3, on average. Also notice that spot number 1, the SLD scanning spot, is off at regular intervals. The only other criterion is that the motion of the spot not exceed a low threshold. It is possible for glare to appear periodically with the landmark spot; in this case it may be confused for the landmark. This case has not been observed, and in any case it is not important because detecting the SLD scanning spot is far more important than detecting the landmark.

**Continued tracking**

Once the SLD spots are identified, the spot tracking changes to a predictive tracker. The SLD landmark location is known. Moreover, one can predict whether it will appear based on its last appearances. In particular, if it has appeared for two consecutive fields, it will certainly not appear in the next two fields. If it has not appeared for two consecutive fields, there is a small chance it will appear in the next (the off time is occasionally two fields), and if it has not appeared for three consecutive fields, then it will certainly appear in the next. If a spot is detected within 20 pixels of the predicted landmark location in a field the landmark is predicted to appear, it will be considered to be the landmark. Otherwise, no landmark will be detected.

The SLD scanning spot is tracked using the assumption that it travels in a circle. The best fit circle is estimated from the spot locations in the initial twelve frames using the modified RANSAC [84] approach described previously. Since we know that the scanning spot travels a fixed radial displacement from field to field, its location can be predicted from its previous location and the best fit circle. This constant radial displacement seems to occur even in fields when the SLD scanning spot is not visible,
and so one must consider its previous location as well as how many fields ago that was. Again, a bright spot must be detected within 20 pixels of the predicted scanning spot location for it to be considered a positive detection. This distance allowed for error in the detected spot center as well as the best fit circle.

Both tracking routines are adaptable. The OCT operator may make adjustments to the scan location or radius while a scan is being acquired. To compensate for this, the best fit circle is recalculated for each field; so long as the changes are sufficiently gradual, they will be followed. The changes may not be gradual, however; for example, the user may stop the scan to reposition it, the user may save a given scan, or the patient may close his eyes or pull away from the OCT. In all of these cases, there will no longer be SLD spots visible in the image that exhibit the expected behaviors. If the SLD spots are not found in their expected locations at their expected times for more than eight fields, then the algorithm presumes that it has lost the spots and reinitializes the entire process again, starting again with twelve fields to gather information.

9.3.4 Measuring Motion – Detecting Retinal Structures

Here we describe the subsystem for following the motion of the optic nerve head in the video image. During both the adaptive and operational phases, retinal structures such as the optic nerve head boundary and the vessels are found. During the adaptive phase, however, the optic nerve head is directly located as well, and a model of retinal features relative to the nerve head center is constructed. During the operational phase, the locations of features in each video field $I$ are compared to the model to infer the
location of the optic nerve head. The next few sections describe how the retinal features are detected, and are common to both the *adaptive* and operational phases.

**Removing the SLD spots and glare**

As noted before, the SLD spots and glare saturate the OCT video camera and so their pixels in the image $I$ can be up to three times brighter than the average intensity of the surrounding retina. The brightness of the SLD spots is actually problematic and interferes with the segmentation of the blood vessels and with the use of the eigenimage techniques for detecting the optic nerve head. So the SLD spots are always removed from the video field image $I$.

To remove each bright spot, bilinear interpolation is used in the surrounding region. In particular, we assume a spot can exert effects in a 20 pixel square around the spot center. This size was chosen empirically and was intended to ensure that no glare effects remained. We wished to err on the side of excessive suppression because of the highly deleterious effects of the SLD spot, when present. For each pixel $(r_i, c_j)$, a bilinear approximation $\hat{I}(r_i, c_j)$ for the image intensity $I(r_i, c_j)$ can be calculated from the pixels in each of the four cardinal directions around it. In particular, for a given image grid

\[
\begin{array}{ccc}
I(r_1, c_3) \\
\vdots \\
I(r_3, c_1) & \ldots & I(r_3, c_9) & \ldots & I(r_3, c_3) \\
\vdots \\
I(r_5, c_3) \\
\end{array}
\]  

(9.7)

the bilinear approximation for $I(r_3, c_3)$ is given by
Figure 9.23: Here are some sample fields showing how the SLD spot can be masked off. In each column, the top image is without the masking, the middle image is using the bilinear transform without the addition of noise, and the bottom image is with the addition of noise.

The bilinear masking, however, is not completely satisfactory. In particular, the 20 pixel region border may have some residual glare, and so the resulting masked region appears to be a locally bright, smooth square. This square can give a strong response.
to the eigenspace recognition techniques. Also, the bilinear surface can have a pixel intensity which varies in a smooth, ramp-like fashion from one side of the square to the other. This generates organized responses from the filter used in the next step to detect retinal structures. These responses, though weak, can be confused with the responses from real structures. Both effects can be addressed by "obscurring" the masked area with multiplicative noise. Additive noise could have been used; because we are only using a small amount anyway, it would not have made a difference and may be more sound, mathematically. Nonetheless, the multiplicative noise was used because it gave an aesthetically more pleasing result. For a given pixel intensity $I$ within the masked area, we compute a new image intensity $\hat{I}$ as

$$\hat{I} = I\gamma$$

(9.9)

where $\gamma \sim N(0, 0.05)$. This amount of noise is sufficient to obscure the bilinearly masked region for a human observer as well as reduce the response of the eigenspace recognition techniques. Also, the filter response becomes weak and disorganized. Alternatives could be found to introducing noise over the bilinear mask. For example, all filter responses in the masked region could be ignored. This, however, would not solve the response problem of the eigenspace technique. Simply discounting all masked regions from being considered valid eigenspace recognitions is not a desirable solution either because there is often glare on the nerve head, which is thus partially masked off as well; masked regions truly are the nerve head at times.
Figure 9.24: These are the 1-dimensional filters used to detect blood vessels in the retinal image.

Filtering with the one dimensional matched templates

Each video field $I$ is next filtered to detect the blood vessels and the nerve head boundaries. The detector is a one dimensional matched template that is run down the columns and across the rows. The results of the row and column filtered images are then summed to produce the final filtered image $I_f$. The template was chosen to mimic the one dimensional cross section of a blood vessel after the minimum filtering, and is shown in Figure 9.24. The row filtered image thus responds best to vertical vasculature elements, while the column filtered image responds best to horizontal elements. The size of the vessel template for the row filter was slightly larger than that of the column filter because the dominant vascular structure typically observed was vertically oriented.

The step and pulse responses of the filter are worth considering because they explain why this vessel detector can also be used to detect the nerve head boundaries and why the SLD spot causes problems and must be removed. These responses are
Figure 9.25: Here are the sample step and pulse responses of the 1-D row filter; the responses of the column filter are similar. The pulse used is a 10-unit long, positive pulse; this pulse simulates the SLD spot in the image.

shown below in Figure 9.25. As can be seen, the positive step response is a positive-negative peak pair, separated by a zero-crossing. A square pulse, which is similar to the cross section of a blood vessel, can be thought of as adjacent negative and positive steps. Thus its response is a combination of a positive-negative peak pair with a negative-positive peak pair, so that a large peak is generated in the middle.

The use of a one-dimensional filter offers much faster performance than a two dimensional template such as that used by Chaudhuri. [57] One might argue that a traditional edge detector such as the Sarkar-Boyer or Laplacian of Gaussian detector might yield superior performance. These detectors were tried as well, but the noise and lack of contrast in OCT images causes the detected edges from retinal blood vessels to be not significantly stronger than edges due to retinal pigment variations. Moreover, the vessel boundaries are broken up, and so approaches similar to that of Can et al. which look for paired vessel boundaries exhibiting opposite local image
gradients do not work well either. So the performance of the traditional techniques was not notably better than the 1-D detector.

In Figures 9.26 and 9.27, sample video fields, $I$, and the filtered results, $I_f$ are shown. As can be seen, the filter responds to filamentous structures like blood vessels with a string of local peaks along the central spine of the vessels. A similar response is generated by the boundaries of the optic nerve head because the nerve head, when seen across a row or down a column, looks like a broad step; therefore we see the positive peak of the detector's step response. It helps further that oftentimes there is a small, dark pigment ring around part or all of the nerve head, which then mimics a blood vessel. The filter response to the masked SLD spots is a relatively weak, disorganized region of small peaks.

One can see the benefit of masking the SLD spot by comparing the filter responses with and without it, as in Figure 9.28. Here the same field was filtered after the spot was masked (left) and prior to masking (center and right). The filter response to the SLD spot is very strong because, in a row or column cross section, the SLD spot appears as a very bright pulse. The filter response is a very strong, negative peak surrounded by smaller, positive peaks; because the SLD spot is so bright relative to the surrounding retina, the surrounding positive peaks can be much stronger than the positive peaks corresponding to actual blood vessels.

The filtered image $I_f$ is then thresholded to select regions more likely to be real retinal structures. We note that a histogram of pixel intensities in $I_f$ can be approximated as a probability distribution function, $p \sim N(0, \sigma)$ for some $\sigma$. We estimate $\sigma$ by finding the standard deviation of pixel intensities in $I_f$, and we lessen the chance of choosing noise by selecting pixels above the intensity threshold $\sigma$. This gives us
Figure 9.26: Here are some sample fields showing the result of the 1-D vessel filter. There are two video fields from 4 subjects and one from a 5th, and, in each case, we have a row of video fields above a row of filter results.
Figure 9.27: Here are some more fields showing the result of the vessel filter.

Figure 9.28: Here is a sample figure showing how the SLD spot can dominate the filter output when it is present. This is why we mask it off. A masked filter result is on the left, and an unmasked filter result is in the middle. One can see that the region of the laser spot can be very negative compared to the rest of the image; a mesh plot of the middle image is on the right so that the relative magnitude of the spot region's response can be appreciated.
only the top 16% of pixels, and this thresholded result is denoted \( I_j \). We see that there is a lot of variation in the noise clutter from field to field, but the dominant vessels and nerve head boundaries tend to remain visible as large, connected shapes. We take advantage of this connected nature of the filter responses from actual retinal structures to create a second level of thresholding. In particular, we threshold the connected, non-zero regions of \( I_j \) according to size. Only those above some minimum size threshold are selected, with the result being denoted \( I'_j \). Examples of \( I_j \) and \( I'_j \) are in Figure 9.29 and 9.30.

The intensity thresholded image \( I_j \) shows both the nerve head and the vessels, but there is much clutter in the image as well. We would like to find structures and have confidence that they are real; for this \( I'_j \) is preferred. However, this specificity comes at a cost; much structure, including the nerve head boundary, is often lost in the formation of \( I'_j \). Hence \( I'_j \) will be more suited to finding the nerve head.

**Finding structures within the filtered image using peak detection**

Once the SLD spot is masked off, we find all peaks in \( I_j \) and \( I'_j \) and then connect them to find the contours of the retinal vessels and the optic nerve head boundaries. Like the initial filtering, the peak detection is performed using a one dimensional search along the image rows and columns. Though more sophisticated techniques exist, this one dimensional technique was fast and gave us satisfactory performance. The assumptions used in this approach are, first of all, that the filter responses to the vessels and nerve head boundaries give local maxima when viewed in row or column cross section. This seems intuitive when the vessels are perfectly vertical or horizontal, but is also the case when they are diagonal. We also assume that the vessel responses are locally, at least, larger than those of the surrounding retina. Finally, we assume
Figure 9.29: This shows how thresholding the video images preserves the vascular structure while mitigating the noise. The middle image is the intensity thresholded image, $I^I$. The rightmost column shows the size thresholded image, $I^S$. 
Figure 9.30: This shows additional examples of intensity and size thresholding.
that the vessels are separated by some minimum distance. Though arterial-venous pairs do often run side by side, we are, in most cases, unable to see them both due to the low contrast. Moreover, we generally only see the largest vessel branches, so that we do not have to worry about vessel cross-overs.

The peak detector runs through each row and column, examining 40 pixel windows. The windows are staggered every half width so that any peak will lie in at least one window. This half width of twenty pixels becomes the minimum peak separation that can be detected. Within a window, all peaks are detected and the largest is selected. To lessen the noise response, the peak amplitude, measured as the difference between the maximum peak value and the minimum intensity within the window, must exceed an empirical threshold. The result of the peak detection \( I^1_j \) is denoted \( I^1_p \), and similarly, \( I^2_j \) corresponds to \( I^2_p \).

For the video fields being used as examples in Figures 9.31 and 9.32 we have the detected peaks displayed for \( I^1_p \) and \( I^2_p \). The peak detections in \( I^1_p \) and \( I^2_p \) are very noisy for several reasons. First of all, the filter response to blood vessels is not a nice clean ridge, but rather a line of peaks akin to a mountain range. This is partly due to the summing of the horizontal and vertical responses, as well as from the fact that the images of the vessels themselves are often broken despite the minimum filtering. Thus the peak detection is broken up rather than lying along a continuous line. Another reason for the noise is that summing the vertical and horizontal peak detections tends to form crosses at peak locations rather than single points. Finally, the \( I^j_i \) themselves are quite noisy, despite the intensity thresholding, and so there are many spurious peaks for the peak detector to find.
Figure 9.31: This shows the result of peak detection in $I^f_1$ (left column) and $I^f_7$, yielding $I^p_1$ (middle column) and $I^p_7$ (rightmost column)
Figure 9.32: Here are more examples
Finding contours among the peaks

We now organize the peaks in $I_p^2$ into contours representing the nerve head boundaries and the vessels, using a coarse organization scheme. The coarseness came from the fact that we don't want to simply chain code the individual pixels in an eight or four-connected way. This is because the peak detector responses to real image structures, such as vessels and nerve head boundaries, are series of separated segments and "x" shapes. Thus an organizer must be able to recognize these as members of a coherent, contiguous structure and act on a coarser representation of the vessels than one-pixel wide contours.

To accomplish this, the organizer works in blocks. On its first iteration, the organizer selects a non-zero pixel in $I_p^2$, with the presumption that this point may be in the middle of an image structure. This point location is taken to be the first point of the first contour, $\xi_{1,1}$. We then determine the most likely vessel orientation around $\xi_{1,1}$, and pick the next contour point, $\xi_{1,2}$ some distance away in that direction. The structures of interest are thin curves, and so each direction is scored by how many non-zero pixels lie within a narrow line in that direction. In particular, the local 5x5 neighborhood of $I_p^2$ about $\xi_{1,1}$ is examined by multiplication with a series of sixteen 5x5, directional matched templates; these are shown in Figure 9.33. Each template, $T_k$ is shaped like a fat line segment extending in the direction $\theta_k = k\frac{\pi}{8}$ from the template center, and the segment's intersection with the template boundary is that direction's candidate for the new point on the contour. After the template multiplications, the $\theta_k$ with the highest scoring $T_k$ wins and its candidate for the new contour point becomes the new end point, $\xi_{1,2}$. Note that $\theta_k$ that were diagonal to the vertical or horizontal axes had an advantage in the voting scheme simply by virtue of their corresponding
Figure 9.33: This shows the directional templates used in the organizer. On the left are the templates used to vote for new growth directions. On the right are the templates used to zero out points in $I_p^2$ after voting has occurred. In these binary template images, black equals 1 and white equals 0.

$T_k$ having more pixels. We were aware of this fact but did not remedy it as it caused no detriment to performance.

The process for contour growth is thus repeated for each $\xi_{1,i}$ to determine $\xi_{1,i+1}$. There are, however, some additional complications which must be addressed. First of all, we need to ensure that we don't count the non-zero pixels in $I_p^2$ more than once while growing vessels. Thus, when a new end point, $\xi_{1,i+1}$ and direction for contour growth, $\theta_k$ are chosen, the nonzero pixels of $I_p^2$ that voted for $\theta_k$ are set to zero. Moreover, a coarse chain coding of the vascular structure is sufficient, and so we also remove all nonzero pixels that voted for $\theta_k \pm 90^\circ$. This zeroing out of pixels is accomplished efficiently by multiplication of the local 5x5 region around $\xi_{1,i}$ with one of a set of 5x5 templates, $T_k^x$, that correspond to the directional templates, $T_k$.

These 5x5 templates are divided by a line perpendicular to $\theta_k$ and are zero on the $\theta_k$ side and one on the $\theta_k + 180^\circ$ side, and are shown in 9.33. Contour growth continues
until no candidate direction around the contour end point gets any votes using the template voting method. At this point, one of two things happens. Recall that it could be that $\xi_{1,1}$ is in the middle of a retinal structure, and that we simply picked one of two possible directions and encoded that portion of the structure to its end. In that case, the rest of the structure remains to be encoded, and we wish it to be encoded as part of $\xi_1$ rather than as a new contour. Thus the organizer goes back to $\xi_{1,1}$ to determine whether any other direction of growth is supported; recall that the zeroing process only removes points within $90^\circ$ of the direction of growth and so points lying more than $90^\circ$ away may allow further growth. If another direction of growth is supported, then we add this point to the beginning of the contour and it becomes $\xi_{1,1}$; not that this requires re-indexing all the previously coded points in $\xi_1$. When growth at the beginning of $\xi_1$ is no longer supported, $\xi_1$ is considered complete and a new point is chosen from the remaining non-zero pixels in $I_p^2$. This new pixel becomes the first point of the second contour, denoted $\xi_{2,1}$, and the whole procedure begins again with $\xi_2$. New contours are created until all pixels in $I_p^2$ are set to zero. Isolated non-zero pixels that don’t support growth in any direction are suppressed, and are not counted in the final tally of contours.

The organizer finds both the blood vessels and the nerve head boundaries in $I_p^2$. Extraneous information can be found as well, however. These spurious contours tend to be smaller and can mostly be eliminated by choosing only contours that are larger than some threshold. The empirically chosen threshold was 5 points. The set of thresholded contours is called $\{\xi_j\}$. Each individual contour, $\xi_j$ is composed of a sequence of points, $\xi_{j_1}, \cdots, \xi_{j_{n_j}}$, where $n_j$ is the number of points in the $j^{th}$ contour.
The contours are encoded in two ways. The first is as the lists of points in each contour, \( \{(\xi_{i,1}, \cdots, \xi_{i,n_i})\} \). The second is as a map of contour locations, \( I_v \), in the image, \( I \). This map is created as the \( \xi_j \) are created. For each \( \xi_{j,i} \), the corresponding pixel in \( I_v \) is set to 1, as are the pixels in the region between \( \xi_{j,i-1} \) and \( \xi_{j,i} \). The particular pixels that are set to 1 are those corresponding to the non-zero points in the template, \( T_k \) that represented the contour growth from \( \xi_{j,i-1} \) to \( \xi_{j,i} \). Samples of the final \( \xi_j \), as represented by \( I_v \), are shown in Figures 9.34 and 9.35.

9.3.5 Measuring Motion – Finding the Nerve Head and Mapping of Retinal Structures

During the adaptive phase only, we locate the optic nerve head during each video field to build a model of retinal structures relative to it. As previously stated, we use a series of three techniques to narrow our choice to one location. Each of the techniques is prone to error, but their sets of false positives are different. The techniques are strategically arranged to take advantage of their strengths and ameliorate their weaknesses. The first technique uses a Hough transform on the peaks of the filtered image. The goal is to locate the circle corresponding to the nerve head boundary. Next, a template is taken from the image region surrounding each of these circles and the dual class eigenimage technique is used to decide whether it contains the nerve head. Only the highest scoring circles are then passed to the third technique. The eigenimage technique is slow and so the initial selection of circles greatly limits the search area. However, the eigenimage approach is not good at determining the actual nerve head radius nor at precisely locating the circle centers. Also, there are false targets on the retina that confuse it. Thus the final technique analyzes the geometric
Figure 9.34: The leftmost image in each row shows $I^2$ so that we can show the structure that is there to be found. The middle image shows the binary image of detected vessel positions, $I_v$. The right image shows the positions of the detected contours, $\xi_j$ overlaid on top of the original video field for comparison.
Figure 9.35: Here are more examples
relationship between the \( \{ \xi_i \} \) and the candidate circles to determine which is most likely to be the optic nerve head.

**The Hough transform**

Here we implemented a Hough transform to find circles in the image of \( I_p \). We don't use \( I_f^2 \) because the nerve head boundaries are often weaker detections than the vessels and the size thresholding completely removes them. Even the intensity thresholding used for \( I_f \) and thus for \( I_p \) doesn't allow consistent detections, but the results we get are sufficient for the Hough transform. Moreover, using a lower threshold admits too much noise and thus slows subsequent steps by allowing too many candidate circles to be found.

For the binary image, \( I_p \in \mathbb{R}^2 \), the Hough transform is a voting scheme that allows every point in \( I_p \) to vote for a set of points in a voting space, \( H \in \mathbb{R}^3 \), that represents all possible circles that point could belong to. The mapping from \( \mathbb{R}^2 \) to \( \mathbb{R}^3 \) is not a function since a given \( (r_i, c_i) \in I_p \) votes for a set of \( (r_h, c_h, \rho_h) \in H \). This set depends on the radius \( \rho \) of the circle we are considering; for a given \( \rho \), \( (r_i, c_i) \) can belong to any circle whose center is distance \( \rho \) from \( (r_i, c_i) \). That is, the locus or possible circle centers to which \( (r_i, c_i) \) can belong to is a circle of radius \( \rho \) around \( (r_i, c_i) \). Formally, the selection of points to vote for is

\[
(r_i, c_i) \rightarrow \{(r_h, c_h, \rho_h)|r_h^2 + c_h^2 = \rho^2 \forall \rho > 0\} \tag{9.10}
\]

so that when all possible radii are considered, the locus in a cone in \( \mathbb{R}^3 \). When a real circle exists, their loci of votes all intersect at the corresponding location in \( H \), and so that location receives the maximum number of votes.
For practical implementation, we made several simplifications. First of all, the range of possible nerve head radii, $\rho$ is limited and so we restricted our section of $\mathbb{R}^3$ to this range. We chose a range from 27 to 48 pixel radii. This is the imaged size after the individual fields are separated and the column resolution reduced by two; the corresponding radii in the Hough space, $\rho_h$ correspond to 8 and 16 pixels, respectively. Digital images are discretized; we further discretized the search range in $\rho$ to increments of 3 pixels to increase the search speed. Furthermore, we also discretized the search range in $r_h$ and $c_h$ by reducing the resolution $H$ by a factor of three. Part of this was to increase speed, and part was to make peak location more likely. As stated before, the detected peaks for the vessels and nerve head boundaries do not form smooth, continuous contours, but rather disjoint series of points, short line segments, and intersecting line segments. Thus reducing the resolution of the Hough voting space increased the likelihood that points corresponding to the same circle would vote for the same bins in the Hough voting space. This fundamentally limits the spatial resolution of circle detection so that a point in Hough space represents a 3x3 block in the video field. As a result of both discretizations, we know the center location and radius of a circle to within three pixels. However, this is sufficient accuracy for the adaptive phase.

We also took advantage of the fact that many of the points in $I_p$ lie within a contour in $I_v$, whose local direction we thus know. This information comes from the $\xi_i$ and the direction of the vector between each pair, $\xi_{i,i}$ and $\xi_{i,i+1}$. This direction corresponds to the direction of the template, $T_k$ that was used to fill in the points between $\xi_{i,i}$ and $\xi_{i,i+1}$ in $I_v$. In particular, say we know the local contour orientation at some $(r_o, c_o)$ to be $\theta \pm \delta$, where $\theta$ is the contour direction and $\delta$ denotes the uncertainty in $\theta$. Then
Figure 9.36: This shows the templates used in the Hough transform. To the left is the template used for the $\rho_h = 17$ Hough space, and to the right are the templates used for $\rho_h = 33$ Hough space. In the bottom right of each figure is the template used when no directional information was available; this template is a full circle.

$(r_o, c_o)$ cannot be part of a circle whose center, relative to $(r_o, c_o)$, is located in the angular intervals $(\theta - \frac{\pi}{2} + \delta, \theta + \frac{\pi}{2} - \delta)$ and $(\theta + \frac{\pi}{2} + \delta, \theta - \frac{\pi}{2} - \delta)$. Conceptually, this means, for example, that a point that is on a vertically oriented contour must belong to a circle whose center is to the left or right, not one that is above or below. We took $\delta$ to be $\frac{\pi}{16}$ so that we reduced the locus of voting points by about $\frac{1}{3}$. This choice of $\delta$ is very conservative, but this was appropriate because of the very coarse directional resolution of the organizer. Figure 9.36 shows templates representing the regions of Hough space that were voted for for given contour orientation. We considered the orientation information to be reliable only in the case of large, $\xi_j$; contours had to have at least 5 points before we had sufficient confidence in their orientation information. But this is the same threshold used to choose the final set of $\xi_j$.

The use of directional information to limit the Hough space voting is faster, but, more importantly, removes some clutter from the Hough space. The effect is to give
Figure 9.37: This shows samples of $H$ for the $\rho_h = 17$ Hough space on the left and the $\rho_h = 33$ Hough space on the right. In the top row, the directional information is used, and in the bottom row it is not. Though the difference is not dramatic, one can see that the top row is somewhat less cluttered, with more contrast between the peak regions and their surroundings.

locations corresponding to actual circles more distinct maxima, and can be seen in Figure 9.37. Hough spaces are shown for two different $\rho_h$ for the same video field, using the directional information (top row) and not using it (bottom row). The darkest points in the image represent those with the most votes. The effect is to give the regions with the most votes more contrast with their surroundings; the effect is subtle, but nonetheless beneficial.
Once voting was finished for the subspaces representing the different radii, we searched each subspace for local peaks. The voting subspaces tend to have many local maxima, any one of which can be the nerve head. We select these maxima in the same way used to select the SLD spots. We find the brightest 100 points in a given Hough space that are at least 2.5 standard deviations above the mean. The number 100 was chosen empirically, and the value of 2.5 standard deviations increases the likelihood that these points are not spurious. Note that doing this in two steps is equivalent to simply choosing the top 100 points; the two step approach was chosen to make it conceptually clearer. In Figure 9.38 we show the same Hough spaces in 9.37 (using the directional information) after the thresholding to 2.5 standard deviations (top row) and the selection of the brightest 100 points bottom row).

The clustering algorithm then finds the clusters within these 100 points. We assume a minimum cluster radius of 15 pixels within the Hough space, $H$; this translates to 45 pixels within the image field, $I$. Also, instead of finding the center of gravity of the cluster, as with the SLD spot, we find the highest value pixel within the cluster, since this represents the circle which received the most votes. Note that the clusters can overlap (their minimum separation for their centers are located less than one radii apart, and so the step of finding the brightest points within each cluster can cause the final centers to be spaced less than 15 pixels apart.)

Thus we end up with a set, \( \{(r_i, c_i)\} \), of "best circles" for each radius, where each circle can be graded by the value of its center in the voting space, $H$. However, it should be noted as well that a large circle with only a fraction of its circumference represented in the image can still have more votes than a small circle which is completely filled in, simply by virtue of having more pixels. Thus the grade of each circle
Figure 9.38: This shows samples of $H$ for the $\rho_h = 17$ Hough space on the left and the $\rho_h = 33$ Hough space on the right. In the top row, an intensity threshold has been applied so that only the points greater than 2.5 standard deviations are used, and in the bottom row, the top 100 of these points are chosen.
is divided by its radius; the resulting, normalized grade represents the average number of pixels per unit circumference. The combined set of best circles for all radii can then be written \( \{(r_i, c_i, \rho_i)\} \). Some sample images showing all the detected circles are shown below in Figures 9.39 and 9.40. Note that only the 40 strongest circles from each video field are shown in the figures just to make them intelligible to the viewer; normally there were 60-80 circles found per video field. As can be seen, the technique is very noisy and finds many false circles. Some of the false alarms are caused by the rampant noise, others are caused by vascular structure being confused with nerve head boundaries. Because there is a lot of structure in the vicinity of the nerve head, there are, oftentimes, several circles of different sizes close to the nerve head. Finally, the nerve head boundaries sometimes produced weak detections in \( I_f \) and thus \( I_p \), and so the Hough transform would miss the nerve head in these cases.

To lessen the instances in which this occurred, the Hough transform would always choose the strongest circle in \( H \) that was within two pixels of the nerve head's location in the last video field. This idea worked because the final determination of the nerve head location was usually right, and because the nerve head usually kept the same position in \( I \) for many video fields at a time. The same phenomena that caused weak boundary detection of the nerve head circle could also cause weak identification by the eigenimage technique, and so the circle closest to the previous detection was always sent to the final, geometric evaluation. This step ensured that the previous location was at least considered by the geometric evaluation. In the case of genuine motion, the geometric technique could often be relied upon to reject this circle. All other circles that are found by the Hough transform are then sent on to the next stage for confirmation using the eigenimage technique.
Figure 9.39: This shows all the circles found by the Hough transform. The left column shows $I_p^I$ to illustrate the available structure. The middle column shows the detected circles overlaid on $I_p^I$, while the right shows them on the video field.
Figure 9.40: Here are more examples
Eigenimage recognition

We use the eigenspaces technique to evaluate each of the circles found by the Hough transform. In particular, we take an 86 x 86 pixel subregion, $T^*(r_i, c_i, p_i)$ that is centered on each circle center, $(r_i, c_i)$, within $I$; this test subimage is reduced in resolution by two and the resulting 43 x 43 subimage is projected into $\hat{\Phi}_N$ and $\hat{\Phi}_R$ to calculate $\eta$. The test subimage depend only on the circle centers and are large enough the accommodate the maximum expected $\rho_i$. However, it is possible that the circles found by the Hough transform are not centered around the actual nerve head center. One reason for this is the large numbers of spurious filter responses, so that a compelling circle hypothesis exists using only a portion of the actual nerve head boundary, with the rest of the circle being composed of noise. Another consideration is the fact that the training subimages used to form the eigenspaces $\hat{\Phi}_N$ and $\hat{\Phi}_R$ were hand picked; thus they were not all necessarily centered exactly on the nerve head. Hence the maximal eigenspace response may not necessarily be at the nerve head center. The dual eigenspace technique described in the Theory and Models section of this chapter has a very gradual, slow response that is very forgiving of displacements of a few pixels. However, to increase the likelihood that the correct $(r_i, c_i, \rho_i)$ would be given the highest $\eta$ score, we also graded $T^*$ displaced a few pixels from each $(r_i, c_i, \rho_i)$. If we define $f_\eta$ as the function which gives the grade, $\eta$ for a given $T^*$, then

$$\eta_i = \max_{j, k} (f_\eta(T^*(r_i + j, c_i + k, \rho_i)))$$

(9.11)

where $-2 \leq i, j \leq 2$. We thus choose the $(r_i, c_i, \rho_i)$ with the largest $\eta_i$ to consider in the final stage. We empirically decided to choose only the top five $(r_i, c_i, \rho_i)$. To improve system speed and reduce spurious responses, we further restricted the choices.
First of all, we accepted only $\eta > 0.8$; this prevented completely spurious responses from going to the next stage simply by virtue of being among the top five circles. Even though the set of eigenvectors used to form $\hat{\Phi}_N$ and $\hat{\Phi}_R$ was chosen to maximize the number of nerve head training images generating a response greater than one, the nerve head in a small percentage of training templates and a corresponding percentage of video frames did not score $\eta > 1$. The reasons for this were not clear; to prevent the nerve head from being rejected because of a low score, we lowered the threshold to 0.9. We could do this because the final stage for the nerve head identification, the geometric relations stage, is good at rejecting the spurious responses that are allowed through because of this low threshold. To further reject spurious responses, we found the maximum $\eta_i$ and only accepted circles whose $\eta_i$ was greater than 0.7 the maximum. Thus in the case where the nerve head responds very strongly with a large $\eta$, the minimum $\eta$ threshold for acceptance will be greater. The circles that were thus chosen were then sent to the next stage. In Figures 9.41 and 9.42 we see the circles that are chosen by the eigenimage evaluation. In the left column of the figures, the video field is shown with the circles from the Hough voting as in Figures 9.39 and 9.40. However, we've superimposed a grid to help clarify which circles are chosen and why. In the middle column, we see the result of calculating $\eta$ for every possible location in the video field. From the figure, one can see how $\eta$ is locally maximum for regions about the nerve head, but that this is not always a global maximum. This is why false positives can result. In the images, the darker regions correspond to larger values of $\eta$. Finally, the rightmost column shows the final circles that are chosen. The circles numbered according to $\eta_i$, with the maximum being labeled "1"; they are also shaded so that the brightest circle represents the maximum $\eta_i$. Recall that there may
be more than 5 because the Hough circle closest to the previous field's identification is always sent for evaluation by the geometric analysis.

**Geometric relations between vessels and the nerve head**

Because of the broad, slow response characteristics of the eigenspace recognition technique, it was not good at precisely locating the nerve head center. Nor did we implement techniques to allow it to discriminate the size of the nerve head. The score \( \eta_i \) only depends on \((r_i, c_i)\) and not \(\rho_i\). Thus a larger or off-center circle might respond more strongly than the correct nerve head circle. We reasoned that the nerve head usually satisfied some abstract geometric relationships with the vessels; the nerve head could ideally be regarded as a central hub from which the vessels emanated radially. In reality, the angles of vessel departure were often not truly radial, or the vessels might bend by some large angle soon after leaving the nerve head. Thus the detected portions of the vessels would not appear to lead to the nerve head center. Vascular bifurcations presented another problem, because only small portions of the vessels might be detected. The segments distal to the bifurcation would usually have orientations that were not radial to the nerve head. Moreover, in some cases, the vessels have a dominant vertical or horizontal orientation, rather than an even radial distribution. Finally, and most importantly, the \( \xi_j \) represented image structures, which included both vessels and nerve head boundaries; there was no easy way to tell from individual \( \xi_j \), which was which. The nerve head boundaries are not radially directed like the vessels, and are, in fact, circumferential.
Figure 9.41: This shows how the eigenimage technique chose circles. The left column shows $I$ with all detected circles superimposed. The middle shows $\eta$ calculated for the entire video field, while the right shows the chosen five circles.
Figure 9.42: Here are more examples
Each candidate \((r_i, c_i, \rho_i)\) from the eigenimage verification stage was then scored according to how well it and the \(\{\xi_j\}\), which represent the detected nerve head boundaries and vessels, fit the abstract geometric model. For each circle, the relative position and orientation of each \(\xi_j\) was considered to contribute a score, \(\nu_{i,j}\) that reflected the conditional probability of that circle being correct, given \(\xi_j\). The scoring system was empirically developed using three of the training subjects, but it worked well with the other three training subjects and is explained below. The final score for \((r_i, c_i, \rho_i)\), denoted \(\nu_i\) was defined as

\[
\nu_i = \pi_{\text{star}}^i \sum_j \nu_{i,j} \tag{9.12}
\]

and the circle with the highest \(\nu_i\) is considered to be the most likely nerve head circle.

The quantity \(\pi_{\text{star}}^i\) is a multiplicative factor that evaluates how well the arrangement of \(\xi_j\) around \((r_i, c_i, \rho_i)\) resembles the idealized model of spokes coming out of a central hub. In particular, \(\pi_{\text{star}}^i\) measures the uniformity of the angular distribution of the \(\xi_j\) about \((r_i, c_i, \rho_i)\).

To calculate \(\nu_{i,j}\), an initial determination is made as to whether \(\xi_j\) is linear or not. Typically, the \(\xi_j\) that belong to vessels (as opposed to the nerve head boundaries) belong to the portions of the largest vessels that are closest to the optic nerve head; these selected vessel segments tend to be straight. Determinations of radial orientation relative to the nerve head only make sense for \(\xi_j\) that are straight lines. Linearity was measured by finding the least squares fit to the points belong to the contour \(\xi_j\) and measuring both the mean squared error and the maximum error to the fit. If either exceeded a certain threshold, the vessel was considered non-linear. For non-linear \(\xi_j\), \(\nu_{i,j}\) depended on whether \(\xi_j\) was contained within the circle \((r_i, c_i, \rho_i)\),
formed a part of the circumference, simply touched the circle boundary, or was some further distance away from \((r_i, c_i)\). The reasoning was that non-linear \(\xi_j\) may be portions of the nerve head boundary, or line contours that incorporated both a vessel segment and a portion of a nerve head boundary. In the latter case, \(\xi_j\) would have a bend of approximately 90 degrees that would make it nonlinear. In either case, the segment should then be close to the perimeter of \((r_i, c_i, \rho_i)\), and, better still, lie along the circumference of \((r_i, c_i, \rho_i)\). However, \(\xi_j\) should not lie completely nor even mostly within \((r_i, c_i, \rho_i)\). This was determined empirically; even though vasculature structure is visible within the nerve head, this structure is usually tangled and does not form clear linear structures for our detector to see. Hence the \(\xi_j\) are rarely from structures within the nerve head. Hence the presence of \(\xi_j\) within a circle suggests that the circle is spurious.

To implement these rules to calculate \(\nu_{i,j}\), the distance was measured between every point in \(\xi_j\) and \((r_i, c_i)\). A point was considered to lie within \((r_i, c_i, \rho_i)\) if it were less than \(\rho - 2.5\) pixels from \((r_i, c_i)\). A point was considered to lie on the perimeter of \((r_i, c_i, \rho_i)\) if it were within \(\rho \pm 2.5\) pixels from \((r_i, c_i)\). Finally, a point was considered to be outside of \((r_i, c_i, \rho_i)\) otherwise. The error range of \(\pm 2.5\) pixels was chosen because of the coarseness of the chain coding and the Hough transform localization. Based on these classifications for individual points, the contour \(\xi_j\) was considered to lie within \((r_i, c_i, \rho_i)\) if more than 80% of its length lay within the circle. The contour was defined as tangential if more than two of its points but fewer than 60% lay along the perimeter. A \(\xi_j\) was considered be circumferential to \((r_i, c_i, \rho_i)\), i.e. to contain a portion of the circle boundary, if more than 60% of its length or at least twelve points lay along the perimeter. Otherwise, \(\xi_j\) was considered to simply touch the
Figure 9.43: Here are some samples of nonlinear contours, $\xi_j$, and their relations to the circles $(r_i, c_i, \rho_i)$. On the left is a contour that doesn't touch the circle, in the middle is one that touches, and on the right the contour is tangential to the circle.

circle if any point lay along or within the perimeter. In Figure 9.43 are samples of a non-linear $\xi_j$ that does not touch $(r_i, c_i, \rho_i)$ (left), that just touches it (middle), and that is sufficiently close to be considered tangential but not circumferential.

If $\xi_j$ was circumferential to $(r_i, c_i, \rho_i)$, then $\nu_{ij} = 0.9$. If $\xi_j$ was tangential instead, then $\nu_{ij} = 0.5$, and if $\xi_j$ is within the circle, then $\nu_{ij} = 0.4$. Finally, for $\xi_j$ that simply touch $(r_i, c_i, \rho_i)$, $\nu_{ij} = 0.75$ minus 0.06 for each point of $\xi_j$ lying within the circle. This scheme was intended to devalue $\xi_j$ that were contained within circles. Again, this scoring scheme was empirically derived based on observations from three subjects.

The rules are more complicated for linear $\xi_j$. In this case, the assumption is that the linear segment may be either a portion of a blood vessel or possibly a large segment of the optic nerve head margin, but we no longer have the case of $\xi_j$ incorporating both. A vessel segment, ideally, should be radially oriented from the nerve head center, and should also come close to, but not within the perimeter of the optic nerve. A segment of the nerve head boundary, on the other hand, should lie over the perimeter of the nerve head. As a final condition, we noted that many spurious
circles were generated that were tangential to the middle portions of $\xi_j$ that were parts of large vessel segments. The reason for this is that all the points on these $\xi_j$ vote for a set of circles in the Hough voting space, $H$, that are displaced from each other by one or two pixels. All of these circles add constructively to form ridges in $H$ that might be misinterpreted as peaks by the clustering algorithm that. We wished to prevent these circles from being chosen. Thus, the first step to calculating $\nu_{ij}$ was to determine whether $\xi_j$ was circumferential to $(r_i, c_i, \rho_i)$, using the same definition as before. If so, then $\xi_j$ was considered to be a part of the nerve head boundary, and $\nu_{ij} = 1$. Samples of circumferential linear $\xi_j$ are in Figure 9.44. In the figure, the contours are shown as before, as white lines with the individual points of $\xi_j$ shown as black dots. The best fit line for $\xi_j$ is shown as a dotted gray line extending out from $\xi_j$. The perpendicular from $(r_i, c_i)$ to the best fit line is shown as a thin black line, while the circle $(r_i, c_i, \rho_i)$ is shown as a black ring.

Otherwise, $\xi_j$ contributed a score, $\nu_{ij}$ that was the product of three numbers: a radially measure, $\pi^+_\text{rad}$ that evaluated how well $\xi_j$ was aimed to the center of the circle $(r_i, c_i, \rho_i)$; a proximity measure, $\pi^+_\text{prox}$, that measured how close the points on $\xi_j$
were to the circle; and a centricity measure, \( \pi_{\text{cent}}^{ij} \), that determined whether \((r_i, c_i, \rho_i)\) was tangent to the central portion of \(\xi_j\). The \(\pi_{\text{rad}}^{ij}\) was scored as the shortest distance between \((r_i, c_i)\) and the best fit line of \(\xi_j\). Next, the \(\pi_{\text{prox}}^{ij}\) was calculated by finding the distance between all points in \(\xi_j\) and \((r_i, c_i)\). Finally, the \(\pi_{\text{cent}}^{ij}\) was measured by determining which point in \(\xi_j\) was closest to \((r_i, c_i)\), and, in particular, how far this closest point was from the nearest endpoint of \(\xi_j\). Hence

\[
\nu_{i,j} = \pi_{\text{rad}}^{ij} \pi_{\text{prox}}^{ij} \pi_{\text{cent}}^{ij}
\] (9.13)

The scoring system, then, was as follows. If the best fit line of \(\xi_j\) passed within 0.5*\(\rho_i\) of \((r_i, c_i)\), then \(\pi_{\text{rad}}^{ij} = 1.0\). If the distance was less than 1.2*\(\rho_i\), then \(\pi_{\text{rad}}^{ij} = 0.8\). Otherwise, \(\pi_{\text{rad}}^{ij} = 0.8 \exp\left(-\frac{(d - 1.2\rho_i)}{4}\right)\), where \(d\) is the distance between the best fit line and \((r_i, c_i)\). Samples of radially oriented \(\xi_j\) that pass within \(\rho_i\) are shown in Figure 9.45. In these cases, the radial orientation is because the circle shown is the nerve head and the \(\xi_j\) are vessels. When incorrect \((r_i, c_i, \rho_i)\) or spurious \(\xi_j\) were chosen, the lines were often not radially oriented, as in Figure 9.47. However, this was not always the case; either a contour or \((r_i, c_i, \rho_i)\) could be spurious, yet the contour would be radially oriented. An example of this is in Figure 9.46 on the left. In this case, the circle is spurious, and so most of the \(\xi_j\) are not aimed towards its center. However, an \(\xi\) representing the actual nerve head boundary is aimed directly at the center of the spurious circle. Another exception to our model of vascular distribution is in the middle of Figure 9.46 where we see that the detected \(\xi_j\), even when they are portions of actual vasculature, may not be radially oriented to the center of the nerve head. Finally, this middle image, as well as the right image in the figure show...
Figure 9.45: Here are some samples of linear contours that are aimed to the center of the circle.

Figure 9.46: Here are some samples of linear contours that are not aimed to the center of the circle.

another exception; vasculature is not always aimed at the center of the nerve head circle, but may be aimed tangentially to it.

In calculating $\pi_{\text{prox}}^{i,j}$, $\xi_j$ were rewarded that just touched the nerve head perimeter, without going within. Thus $\pi_{\text{prox}}^{i,j} = 0.5$ for $\xi_j$ that lay within $(r_i, c_i, \rho_i)$, where within is defined as before. Otherwise, for $\xi_j$ that at least touch the perimeter of $(r_i, c_i, \rho_i)$, then $\pi_{\text{prox}}^{i,j} = 1 - n$, where $n$ is the number of points of $\xi_j$ within $(r_i, c_i, \rho_i)$. Otherwise, $\pi_{\text{prox}}^{i,j}$ decays exponentially, so that $\pi_{\text{prox}}^{i,j} = \exp\left(-(d - 1.2\rho_i)/100\right)$; $d$ is the minimum distance between $\xi_j$ and $(r_i, c_i)$. Typically, as with nonlinear $\xi_j$, the nerve head did not
Figure 9.47: Here are some exceptions to the model. On the left is an example of a linear contour that is aimed to the center of a spurious circle.

Figure 9.48: Here are samples showing that \( \xi_j \) intruding into a circle was usually, but not always, an indication that the circle was not the nerve head.

usually have internal structures that were represented by \( \xi_j \); thus a contour intruding upon a circle was usually an indication that the circle was spurious, as seen in the left and middle images in Figure 9.48. Of course, there were exceptions, as seen in the right image of the figure.

The \( \pi_{\text{cent}}^{ij} \) was calculated as \( \pi_{\text{cent}}^{ij} = \max(0.4, 1 - n) \), where \( n \) is the order along \( \xi_j \) of the closest point of \( \xi_j \) to \((r_i, c_i, \rho_i)\). For the end points of \( \xi_j \), \( n = 0 \), for the next points in, \( n = 1 \), and so on. In the special case where \( \xi_j \) was tangential to \((r_i, c_i, \rho_i)\), (defined as before, so that many of its points touched \((r_i, c_i, \rho_i)\) and not to be confused
Figure 9.49: Here are cases where the long contours corresponding to vessels have generated spurious responses in the Hough space that are tangential to the vessels.

with the best fit line of \( \xi_j \) being tangential to the circle, then \( \pi_{ij}^{proz} \) was adjusted to \( \pi_{ij}^{proz} = -0.75/(\pi_{cent})^2 \). This was done because circles that were tangential to the central portions of \( \xi_j \) representing long vessels often scored well, otherwise, and so we needed to penalize them particularly harshly. Examples of this are in Figure 9.49.

Once all the \( \nu_{ij} \) are calculated, we calculate \( \pi_{star}^i \). The radial distribution of the \( \xi_j \) around \((r_i, c_i, \rho_i)\) is only assessed for the linear \( \xi_j \), whose best fit line came within 0.5\( \rho_i \) of \((r_i, c_i)\). For these \( \xi_j \), the orientation angle, \( \theta_{i,j} \), is defined as the angle of the directed vector from the point of \( \xi_j \) closest to \((r_i, c_i)\) to the point of \( \xi_j \) furthest from \((r_i, c_i)\). A determination is then made of the angle of the smallest sector which contains all the \( \theta_{i,j} \). The final result was \( \pi_{star}^i = 1 + \phi_i \cdot \frac{0.4}{\pi} \). Examples of perfectly even, spoke-like distributions of vasculature were never seen, but when vessels were radially oriented, the correct nerve head circle was usually closer to this model than spurious circles. An example of this is in Figure 9.50. For each circle, only the radially oriented vessels are shown.

Once the \( \nu_{i,j} \) and \( \pi_{star}^i \) were calculated, \( \nu_i \) was calculated. Again, these rules are somewhat convoluted because they were empirically tuned from examination of test
Figure 9.50: Here are a pair of images showing how the correct circle is often closer to the ideal, uniform spoke distribution of vessels. The rightmost image is closer to the ideal than the leftmost; note though that for the leftmost, though the detected vessels are oriented towards the center of the circle, they do not enter through a wide range of angles.

cases. The \((r_i, c_i, \rho_i)\) with the largest \(\nu_i\), denoted \((r^*, c^*)\), was then considered the most likely location of the nerve head.

**Building a structural map:**

During the *adaptive* phase, we used the determinations of the nerve head location, \((r^*, c^*)\) as well as the vasculature location, \(I_v\), to construct a voting space, \(V\), that represented the distribution of the retinal structures around the nerve head. As each video field was analyzed, its data was added to \(V\) so that the accuracy and completeness of \(V\) increased with time. We use the structural representation in \(I_v\) rather than in \(\xi_j\) because the \(\xi_j\) represent only a sampling of points along each structure in \(I\). The representation in \(I_v\), however, represents the most likely locations in \(I\) of the structures represented by \(\xi_j\). The coordinate system for \(V\) was constructed so that the origin represented the center of the nerve head, which was \((r^*, c^*)\) for every video field. The non-zero points in \(I_v\) thus voted for the corresponding points in \(V\), based on their relative positions to \((r^*, c^*)\). Hence a pixel \((r_v, c_v) \in V\) votes for the location
\((r_v, c_v) - (r^*, c^*)\) in \(V\). Note that we have no a priori restrictions for the nerve head position in \(I\). Hence it is possible for \((r^*, c^*)\) to lie along the image borders of \(I\), and so \(V\) has to twice the size of \(I\) in both the row and column directions. Before the first video field is analyzed, \(V\) is a matrix of zeros, since we have, as yet, no information about vessel location. As \(I_p\) is found for each video field, its non-zero locations vote for the corresponding coordinate locations in \(V\). This voting was performed efficiently by adding \(I_p\) to the appropriate subregion of \(V\).

This addition of \(I_p\) to \(V\) continued through the adaptive phase of the nerve head tracking algorithm. For the \(I_p\) to add coherently in \(V\) requires that the determination of \((r^*, c^*)\) be correct. This, in fact, was not always the case, because all stages of the nerve head recognition could have errors. The initial location of circles, \((r_i, c_i, \rho_i)\) by the Hough transform produced many false positives due to the large amount of noise in \(I_p\). Some of these false positives were due to random noise, though some recurred predictably due to the effects of line segments in \(I_p\), as described previously. The eigenimage technique could also produce false positives by having larger values of \(\eta_i\) at incorrect retinal locations. Part of this was because, in some cases, a portion of the retina simply projected more strongly into \(\hat{F}_N\) than did the actual nerve head. However, this effect was limited because that region would have to be selected by the Hough transform first, before it could be scored. Finally, the geometric relationships recognition could also score erroneous circles more strongly than the actual nerve head. This depended highly on the positions of the final circles that reached this recognition stage, as well as the detected \(\xi_j\) for that video field. Both varied greatly from field to field because the large amount of noise in \(I_p\) affected the Hough transform and the existence of the \(\xi_j\).
The displacement, $\Delta$, between the detected nerve head position and the actual position in $I$ tended to be randomly distributed in a uniform manner, as opposed to being systematic. Thus the correct determinations added up in a consistent, coherent fashion from field to field, while the erroneous detections did not. Note that modelling $\Delta$ with a uniform distribution holds true only when one considers data from a sufficient number of fields. When the SLD location is in a certain position, it can occur that the same erroneous location is systematically preferred. It is also theoretically possible for a person to have some retinal feature, such as a histoplasmosis scar, which, because of its high contrast, will always generate a stronger response than the actual nerve head. This is not a problem for creating a map of retinal structure, however. So long as the same retinal structure is consistently located, a map will be constructed relative to that structure, and the scan path will be located relative to that structure. In a worst case scenario, an operator would have to manually locate the nerve head relative to that structure, and the system would calculate scan path relative to the nerve head instead.

In Figures 9.51 and 9.52 samples of the map, $V$ are shown for each of the training subjects. The maps are shown with the data from 10 video fields, and then with the data from 50 video fields.

### 9.3.6 Measuring Motion – Correlating Structures with Our Model

Once the map, $V$, was sufficiently complete, the system left the adaptive phase and entered the operational phase. An empirical determination was made that sixty video fields, or one second's worth of video, was a sufficient adaptive period. At this point, the system stopped finding the nerve head location, $(r^*, c^*)$ and instead
Figure 9.51: Here are examples of how $V$ converges to a coherent model of the vascular distribution. The left images show sample video fields to illustrate the actual vasculature. The middle images show $V$ after 10 video fields, and the rightmost shows $V$ after 50 fields.
estimated it by performing the cross correlation between $I_o$ and $V$. The map $V$ represented the consensus over the adaptive period of retinal structure relative to the nerve head, while $I_o$ represented the structure for a given field. The cross correlation thus finds the best agreement between the individual field and the consensus. This cross correlation, $C = I_o \ast V$ was performed efficiently in the frequency domain. In particular, if $\mathcal{F}(\cdot)$ denotes the two dimensional Fourier transform and $\mathcal{F}^{-1}(\cdot)$ denotes the inverse two dimensional Fourier transform, then

$$C = I_o \ast V = \mathcal{F}^{-1}(\mathcal{F}(I_o)\mathcal{F}(V))$$

The location of the maximum in $C$ then corresponded to the shift in $I_o$ needed to achieve the best agreement with $V$. Thus this technique effectively models the retinal motion as simple translations. However, this model is adequate for the limited field.
of view of the OCT camera and the motions seen during nerve head scans. The peak in C is usually sharp, so that its detection is unambiguous. Some examples can be seen in Figure 9.53. The plots are shown as mesh plots so that the relative elevation of the single peak can be made clear.

In the examples, one notes non-zero values near the far margins of C; these represent large shifts of I_p and one would not expect them to have any correspondence with V. These values are artifacts of the Fourier technique of cross-correlating and come about from the fact that the Fourier cross correlation actually calculates a circular cross-correlation, so that correlation values near one border of C represent shifts that are not far removed from shifts represented by the opposite border of C.

When the nerve head location (r^*, c^*) is inferred from the maximum location in C, the shifted map I_v can be added to V as before. Doing so increases the signal to noise ratio for the feature locations in V that produced the peak in C; these features are presumably correct and so the influence of previous, erroneous locations during the adaptive phase is further reduced. The strength of the correlation can also be used as a measure of the quality of the match. This is important, because in cases where motion occurs, the video image often becomes extremely blurry during the actual saccade. In these blurry video fields, no real features are visible. However, spurious ξ_j may be found due to the usual noise in I_p. These spurious ξ_j will usually have far weaker correlations with V, however, so that the peak in C will be small. An empirical threshold was set at half the value of the most recent correlations. This was done rather than a fixed threshold because each individual will have different structures and will have different correlation values. Also, the map V changes each field as I_v is added. These changes occur slowly, however, after the initial twenty fields,
Figure 9.53: Here are samples of $C$, the correlation between $I_v$ and $V$, to illustrate how sharp the peak usually is.
or so; hence one can look back over the past twenty fields to gauge the value for a typical, correct correlation. Video fields whose correlation falls below this threshold are considered suspect and their information is rejected. This empirical threshold detected many of the bad video frames, though not all. X instances were found of video fields that had a maximum correlation at the wrong location. Increasing the threshold further will not solve this problem, because in one case, at least, the correlation value was not much different from many nearby, correct correlations. The false positive rate is already larger than the true positive rate, so further increases would only make this problem worse.

**Results**

We evaluated the algorithm on video sequences from the six individuals used for training $\Phi_N$ and $\Phi_R$ and from three new subjects. We qualitatively evaluated the training subjects; in all six cases, the algorithm successfully built a model and tracked both the nerve head and SLD beam over sequences of 200-400 fields. The correlation technique detected all fields that were too blurred for correct detection to be performed. For the three new subjects, we performed a quantitative evaluation on the operational phase after a vasculature model was built. We evaluated each subject for approximately 950 video fields; we first noted whether a field was accepted or rejected based on $\chi$, and then determined the error in the detected nerve head location, $\delta_i = ||\nu_i - \nu||$. We had no ground truth for the video fields; however, $\delta_i$ was usually less than 5 pixels, so displacements more than a few pixels were obvious against the visible nerve head features. Finally, when the SLD spots were present, we noted whether they were detected.
<table>
<thead>
<tr>
<th>field status</th>
<th>$\delta_i &lt; 5$</th>
<th>$5 &lt; \delta_i &lt; 10$</th>
<th>$\delta_i &gt; 10$</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>rejected</td>
<td>27</td>
<td>1</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>accepted</td>
<td>2809</td>
<td>1</td>
<td>2</td>
<td>2812</td>
</tr>
<tr>
<td>total</td>
<td>2836</td>
<td>2</td>
<td>4</td>
<td>2852</td>
</tr>
</tbody>
</table>

Table 9.1: The error performance

We inspected 2852 video fields, and the results for $\chi$ and $\delta_i$ are in table 9.1. In the 1996 fields in which the scanning spot was present, it was missed in 142 fields. In the 800 video fields for which the landmark spot was present, the detector missed it in 76.

Low correlation values had two causes. In cases where the fields were blurred, very little vascular structure existed for the 1-D filter to find. Hence there were few $\xi_j$ and so $V_F$ had few non-zero regions to correlate. In the other cases, the nerve head moved to the video field edge. The map $V$ was generally constructed with the disc at the field center, and so moving the nerve head to the field edges brought new vasculature into view that was not modelled in $V$. Thus the only $\xi_j$ that could correlate with $V$ were those nearest to the nerve head. In the two cases of failure, where $\delta_i$ was greater than 10 but $\chi$ was too high for rejection, the nerve head had moved to the screen edge and a spurious correlation peak was larger than the correct peak. The rejection criterion of $50\%$ of the average $\chi$ rejected two frames with low $\delta_i$ for every frame with high $\delta_i$. However, this is acceptable because false rejections occur in less than $1\%$ of video fields. Moreover, of the accepted video fields, $99.93\%$ had $\nu_i$ within 10 pixels of its target, and $99.75\%$ had it within five pixels. Ophthalmologists measure distances in *disc diameters* (DD), using the nerve head diameter as a measuring unit.
The mean nerve head diameter observed was 69 ± 12 pixels (mean ± std dev, range 52-97), and so five pixels corresponds to 0.07 DD. Currently, motions as large as 1-2 DD may go undetected, so this is a vast improvement.

The SLD spots were located in 90% of the fields in which they were present. Many misses occurred because the video field caught the SLD while it was switching between the scanning and landmark locations, and so one spot was very faint. A 90% detection ratio is acceptable because the SLD has smooth motion, and so missed locations can be interpolated. Figure 9.54, shows results from video sequences of different B-scans, most of which featured ocular motion. We show the first video field of the scan, and draw the path of the scanning beam in the video field as a dark line; it is a circle as expected. The path of the scanning beam relative to the nerve head position in this first field is shown as a white line with gray markings where actual SLD detections were made.

Conclusions

We developed a system to find the true OCT scan path in the presence of ocular motion, using the OCT video camera. The system works extremely well, and is able to correctly detect the nerve head position in 99% of all video fields. It is noteworthy that this system works so well using the video footage from the OCT machine, which is not red-free but is instead near-IR sensitive. Hence the contrast is exceptionally poor and we lack many of the vascular features such as bifurcations upon which other algorithms for analyzing retinal images have relied. A single scan requires 1.5 seconds and thus 90 video fields; hence one or two fields may be lost per scan. However, this could easily be made up by obtaining one or two additional scans to fill in the gaps.
Figure 9.54: Here are examples of our final results. In each case, the images are grouped in sets of two rows. In the upper row, the positions of the nerve head through the sequence of video fields is shown; a dot is placed at the nerve head's position in each video field. The positions are superimposed on the first video field of the sequence. In the lower row, we see the actual path the SLD beam travelled, displayed as a white line. On the line are gray dots indicating precisely where the SLD beam was actually detected. All positions are relative to the initial video field, which is displayed as the background. The path the SLD beam took in the video field is shown as a black circle.
Moreover, this is not an onerous requirement, as clinicians currently obtain multiple repeated scans, and this system grants a surety of the scan location that is currently lacking. Our system was intended to be a prototype for developing our theory, and thus it runs off line using previously digitized video. Future research directions include real-time implementation.
CHAPTER 10

CONCLUSIONS

10.1 Current Achievements

This work has explored three main sets of problems. First of all, we have characterized the repeatability of OCT derived measurements of retinal thickness. From the work here, we have found these measurements to be highly repeatable, over many different and clinically relevant situations. If one uses a session of five measurements, and takes the average, then this average will vary less than 10 \( \mu m \) between sessions (99% CI). This is more than good enough for most clinical uses, and so one should only rarely require more than five measurements. Moreover, if one is in a hurry, only one or two should do, with an expected variability of less than 15 \( \mu m \) (99% CI). This too is sufficient for most clinical situations. The variability is the same even when the sessions are separated by up to 18 months, and so following patients over normal clinical follow-up schedules should be alright. We have also found that so long as one obtains a reasonable image, then there is no change in variability with image quality. Hence one need not worry if the scan, as it is acquired, has a different coloration on the Humphrey system than previously acquired scans did. The two sets of scans can be directly compared without adjustment for image quality.
We have also created a retinal measurement system with much better performance than that which comes with the Humphrey system. This system should be able to increase clinical utility of the OCT by providing reliable measurements with fewer errors. Our system is also more usable because its results are exportable, and we have written a GUI to allow correction of erroneous results. Even if the Humphrey system adds such a GUI, however, our system would still be superior because it would require correction less often and thus save time.

Finally, we have developed a prototype system to analyze the fundus video of the Humphrey OCT system and report back the actual scan path. The system works and on 99% of video frames is able to report the nerve head position to a precision of 0.1 DD. Currently, motions as large as 1-2 DD may go undetected, and so this is a large improvement.

10.2 Limitations

All of these projects had some limitations. The repeatability studies were all performed through the fovea on normal subjects. The fovea gives one a very distinguishing feature in the B-scans, so that one can ensure that scans are acquired from the correct location. Surety of location, however, is one of the weaknesses of OCT imaging. In many pathologies, one is interested in measurements away from the fovea, or, in some cases, the fovea no longer provides a clear landmark. The expected variability reported in this work only serve as a minimum bound for expected variability in cases of pathology.

The boundary detector, likewise, was designed for normal subjects. Many pathologies disrupt the normal retinal architecture, and thus the retinal model on which the
boundary detector was built. The video analysis system was designed to analyze nerve head video only, because this is the only portion of the eye with a sufficient number of reliable features. Without a better camera or a wider field of view, tracking of scans over the fovea may be limited. Finally, the system does not run in real time, but rather offline.

10.3 Future Directions

Some suggested directions for future work include the following. First of all, we would like to perform variability studies away from the fovea, or in subjects with macular edema so that there is no longer a clear foveal landmark. We would also suggest redesigning a boundary detector to work with the expected retinal variations found in many conditions that completely disrupt the retinal model, such as subretinal fluid, retinoschisis, macular holes, and epiretinal membranes. Finally, we would like to adjust the video analysis system to run in real time, from digital video.


