Automated Analysis of Normal and Glaucomatous Optic Nerve Head Topography Images

Nicholas V. Swindale, Gordana Stjepanovic, Adeline Chin, and Frederick S. Mikelberg

Purpose. To classify images of optic nerve head (ONH) topography obtained by scanning laser ophthalmoscopy as normal or glaucomatous without prior manual outlining of the optic disc.

Methods. The shape of the ONH was modeled by a smooth two-dimensional surface with a shape described by 10 free parameters. Parameters were adjusted by least-squares fitting to give the best fit of the model to the image. These parameters, plus others derived from the image using the model as a basis, were used to discriminate between normal and abnormal images. The method was tested by applying it to ONH topography images, obtained with the Heidelberg Retina Tomograph, from 100 normal volunteers and 100 patients with glaucomatous visual field damage.

Results. Many of the parameters derived from the fits differed significantly between normal and glaucomatous ONH images. They included the degree of surface curvature of the disc region surrounding the cup, the steepness of the cup walls, the goodness-of-fit of the model to the image in the cup region, and measures of cup width and cup depth. The statistics of the parameters were analyzed and were used to construct a classifier that gave the probability, $P(G)$, that each image came from the glaucoma population. Images were classified as abnormal if $P(G) > 0.5$. The probabilities assigned to each image were in most cases close to 0 (normal) or 1 (abnormal). Eighty-seven percent of the sample was confidently classified with $P(G) < 0.3$ or $P(G) > 0.7$. Within this group, the overall classification accuracy was 92%. The overall accuracy of the method (the mean of sensitivity and specificity, which were similar) in the whole sample was 89%.

Conclusions. ONH images can be classified objectively and dependably by an automated procedure that does not require prior manual outlining of disc boundaries. (*Invest Ophthalmol Vis Sci.* 2000; 41:1730–1742)

Glaucoma is a slow and irreversible neurodegenerative disease, the onset of which is usually not detected by the patient. Diagnosis may be based on a combination of variables,[1,2] but the most dependable single index is probably the identification of a characteristic pattern of visual field defects. However, these defects may only appear after a substantial amount of retinal damage has occurred.[3–6] There is a widely accepted need, therefore, for a method that can reliably detect glaucomatous damage at an early stage so that treatment to prevent further progression can be instigated. Ideally, such a test would have high sensitivity and specificity and be quickly and cheaply administered to large numbers of persons in the normal population, especially those most likely to be at risk of the disease, such as the elderly and those with a family history of glaucoma.

The introduction of the confocal scanning laser ophthalmoscope, such as the Heidelberg Retina Tomograph (HRT; Heidelberg Engineering, Heidelberg, Germany), which is able to obtain accurate three-dimensional images of the surface topography of the optic nerve head (ONH),[7–18] offers a promising means for early detection of glaucoma. A number of studies have shown that morphologic indices calculated from images of the ONH differ significantly between normal eyes and eyes with glaucomatous visual field defects.[19–32] Parameters calculated from combinations of these indices can be used to diagnose the presence of glaucomatous field loss, within the populations from which normative values were obtained, with sensitivities and specificities that are typically in the range of 80% to 90%.

These methods all rely on shape parameters that are calculated by software after an initial stage in which a technician or clinician uses a computer mouse to manually outline the edge of the optic disc. This outlining process has been controversial, because different observers do not always agree where the disc margins should be placed, and this introduces an element of uncontrolled variability into the morphologic analysis.[33] A solution to this problem is to develop automated image processing algorithms that do not require manual intervention. We propose such a method and show that the diagnostic accuracy of the parameters extracted by it is comparable to that of current methods based on prior manual outlining. The technique is based on parametric mathematical modeling of ONH shape, and it works by finding, for each image, those model parameters that produce the greatest degree of similarity between the model and the image. The parameter values are then used as descriptors of ONH morphology and as a basis for further morphologic analysis. To validate the technique, we applied it to a database of 100 images from eyes screened to

From the Department of Ophthalmology, University of British Columbia, Vancouver, Canada.

Supported by the British Columbia Health Research Foundation and the Glaucoma Research Society of Canada.

Submitted for publication June 23, 1999; revised December 1, 1999; accepted January 18, 2000.

Commercial relationships policy: P(NVS, AC); N(GS, FSM).

Corresponding author: Nicholas V. Swindale, Department of Ophthalmology, University of British Columbia, 2550 Willow Street, Vancouver, BC V5Z 3N9, Canada. swindale@interchange.ubc.ca

exclude the presence of glaucoma, and to 100 images from eyes deemed, on the basis of visual field testing, to show early glaucomatous visual field damage.

METHODS

The Model

Images of normal and glaucomatous ONHs obtained with the confocal scanning laser ophthalmoscope typically exhibit a central, roughly circular depression of variable width and depth (the cup), superimposed on a relatively smooth surface with a variable degree of curvature (the rim region). This curvature is typically convex in normal subjects and is caused by the layer of ganglion cell axons becoming, of geometric necessity, increasingly thick as the axons converge toward the optic nerve. These regularities in shape suggest the possibility of description by a mathematical model. Such a model would ideally have a small number of parameters, whose interpretation in anatomic terms would be relatively straightforward. We devised and investigated a model (described in detail in the Appendix) that has 10 free parameters, each of which corresponds to a specific aspect of ONH morphology (Fig. 1). A complete description of these parameters is given in Table 1. A standard nonlinear, least-squares fitting technique[34] was used to adjust the values of the parameters to produce the closest degree of similarity between the model image and the particular ONH image under consideration. Figure 2 shows examples of ONH images and corresponding best-fitting model images.

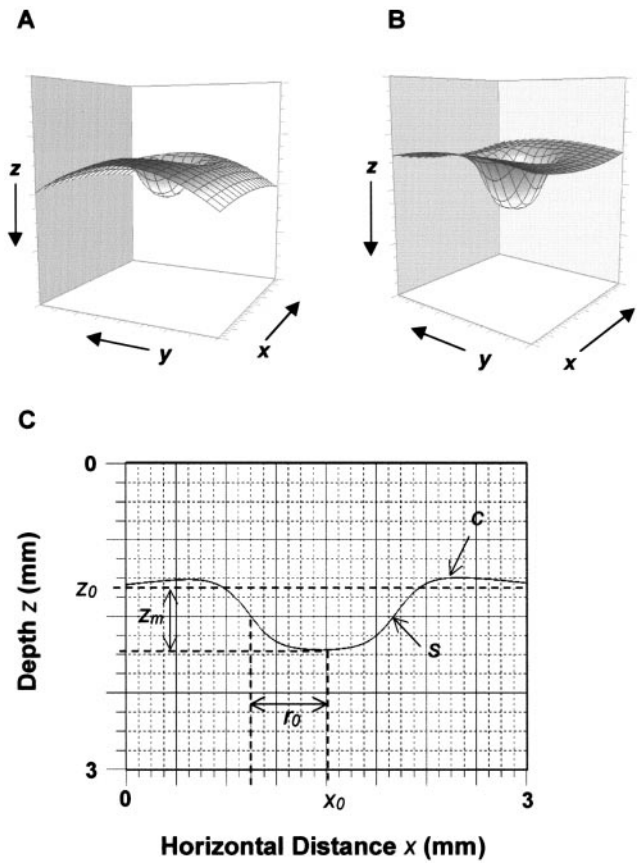

FIGURE 1. (**A, B**) Wire-mesh plots of model ONH profiles based on mean parameter values from (**A**) normal images and (**B**) glaucomatous images. Each of the three axes is approximately 3 mm long. (**C**) One-dimensional profile through a model image along the horizontal axis, at $y = y_0$, illustrating the meaning of some of the model parameters. Parameter c describes the overall curvature of the image in the horizontal axis; s determines the steepness of slope of the cup walls; z_m is a measure of cup depth; r_0 is the distance of the cup wall (at half-height) from the center of the cup at x_0; z_0 is the baseline height of the image (all depth measures are relative to z_0).

TABLE 1. Model Parameters

Parameter Name	Symbol	Description
Nasotemporal slant	a	Overall component of tilt in the nasotemporal axis (mm/mm)
Vertical slant	b	Overall component of tilt along the vertical axis (mm/mm)
Horizontal image curvature	c	Overall curvature along the nasotemporal axis (mm^2/m)
Vertical image curvature	d	Overall curvature in the vertical direction
Cup position	x_0, y_0	Position of center of cup in image (mm)
Cup radius	r_0	Distance from center of cup to the cup wall at half-height (mm)
Cup slope	s	Slope of cup wall (mm)
Cup depth	z_m	Depth of the cup (mm)
Vertical offset	z_0	Offset of the image in the vertical direction (mm)

After these initial fits, a number of morphologic indices were calculated, using the best-fitting model image as a guide. The selection and definition of these was guided by their usefulness in discriminating between normal and glaucomatous images. The methods used to calculate the indices are described in detail in the Appendix, and a summary is given in Table 2. In brief, the model was first used to establish the coordinates of a circular region that covered the cup (Fig. 3). Within this region we calculated (from the real image) 1) an overall measure of the steepness of the cup walls (g_r) and separate temporal (g_r^T) and nasal (g_r^N) components of this measure; 2) a measure of the goodness-of-fit of the model to the image in this region (f_R): This value would be expected to be large in images in which the cup was irregular in shape and not well-described by the model; 3) a measure of the goodness of fit of the image to a model without a cup—that is, a smooth parabolic surface (f_P): This measure would be large in images with large cups and small in images in which a cup was small or absent; and 4) a measure of maximum cup depth (z_{500}) that was the average of the 500 largest depth values present for image pixels within the cup region.

This method was applied to a database of 100 images obtained from eyes screened to exclude the presence of glaucoma and 100 images obtained from eyes with open angles and showing visual field changes indicative of glaucoma. Criteria for subject selection are described in detail in the next section. The model fitting, analyses, and classification[35] were implemented with the aid of a batch-processing language that allowed the calculations to be performed on each of the images automatically without user intervention. A 233-MHz Pentium II computer performed the computations, and the total processing time for each image was approximately 3 to 6 seconds.

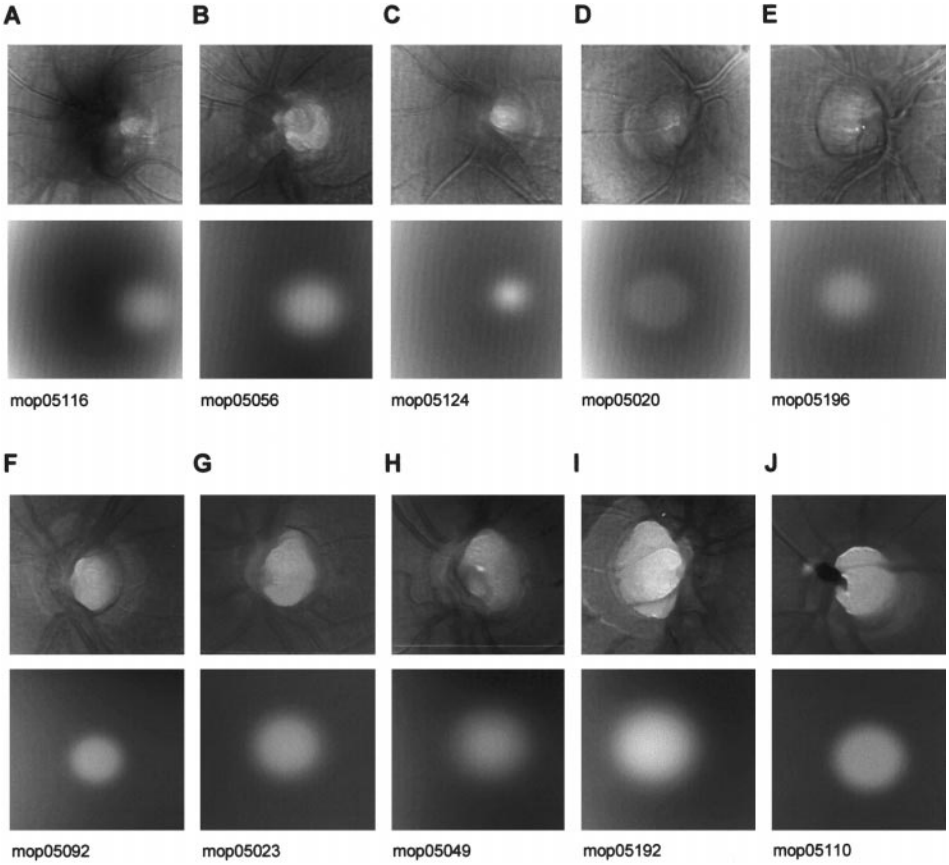

FIGURE 2. (**A** through **E**) Examples of normal ONH images (*top*) together with the corresponding model image (*bottom*) chosen at random from the correctly classified normal population. (**F** through **J**) Image and model pairs, also chosen at random from the correctly classified glaucoma population. Contrasts and brightnesses of image pairs have been adjusted for clarity.

Criteria for Subject Selection

Normal Subjects. The method of obtaining images from eyes in the normal group was as follows:

> Advertisements asking for volunteers were placed in local media and in locations likely to be frequented by the elderly. Staff at the Eye Care Center and their friends and relatives were also recruited. Volunteers were told about the nature and purposes of the study and asked some preliminary questions. They were excluded at this stage if questioning revealed that they had eye disease or a history of eye disease known to be related to glaucoma (e.g., pigmentary dispersion syndrome), if they had a condition such as keratoconus or cataract severe enough to interfere with scanning, if they did not have normal corrected visual acuity, or if they had strabismus (which often causes fixation difficulties during scanning).

If they passed the screening questions, volunteers visited the clinic and were given a consent form to sign. A brief medical history, including details of any relative(s) who had glaucoma was obtained, and the following tests were performed: a Humphrey (San Leandro, CA) visual field test (threshold 30-2) of both eyes, determination of intraocular pressure (IOP) measured in both eyes by tonometry (Tono-Pen XL; Mentor, Santa Barbara, CA), and HRT scans,

TABLE 2. Secondary Parameters

Parameter Name	Symbol	Description
Cup gradient measure	g_r	Overall steepness of cup walls
Cup gradient measure temporal	g_r^{T}	Overall steepness of cup walls on the temporal side
Cup gradient measure nasal	g_r^{N}	Overall steepness of cup walls on the nasal side
Fit in central region	f_{R}	Dissimilarity between the model and image in the cup region
Fit of parabolic function	f_{P}	Dissimilarity between the image and a smooth parabolic surface without a cup
Maximum cup depth	z_{500}	Average of the 500 largest depth values in the cup

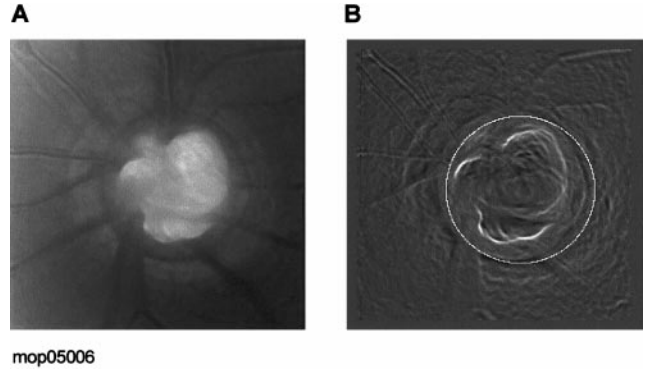

FIGURE 3. (**A**) Topographic image of an ONH; (**B**) the radial gradient map calculated from the same image, with a *circle* enclosing the analysis region *R*.

10° × 10° in size, of each eye through undilated pupils. To obtain these images, at least three separate scans of each ONH were obtained, and a single mean of three images was calculated.[36] When more than three scans were obtained, the set of three giving the lowest SD, as reported by the HRT software, was chosen.

The following criteria were applied for inclusion in the normal group: The visual field had to be within normal limits in both eyes as defined by the Humphrey glaucoma hemifield test (some subjects whose results were borderline on this test were included after further clinical evaluation of their fields), there had to be less than 30% fixation losses during testing, the IOP had to be 21 mm Hg or less, and the SD obtained on averaging three separate HRT scans had to be 50 μm or less. After performing scans and visual field tests on 107 volunteers, two subjects less than 25 years old were excluded on grounds of age, four subjects were excluded because of severe (less than −7 D) myopia, and one subject was excluded because of abnormal visual field test results. A family history of glaucoma was not used as a criterion for exclusion. As it turned out, many (33/100) of the volunteers reported a positive family history, and this was often the reason that subjects gave for volunteering in the first place.

In the majority of cases, scans from both eyes of each subject were available, and only one was chosen for inclusion in the final database. This was done either at random, or in such a way as to make the numbers of left and right eyes equal.

Glaucoma Subjects. Images from glaucoma subjects were selected by first reviewing the files of approximately 415 patients (of FSM) for whom HRT image data were available. The files were reviewed consecutively until images from 100 eyes had been obtained, satisfying the following inclusion criteria: 1) open angles; 2) 20 visual fields (Humphrey or SITA 30-2; Humphrey Instruments) indicative of glaucomatous damage based on the criteria of Mikelberg et al.,[21] i.e., the presence of (a) three adjacent points down by 5 dB with one of the points being down by at least 10 dB, (b) two adjacent points down by 10 dB, or (c) three adjacent points just above or below the nasal horizontal meridian down by 10 dB. None of the points could be edge points except those immediately above or below the horizontal meridian. The field taken closest in time to the HRT scan was used for evaluation. In all except one case, this was within 6 months of examination by the HRT; 3) absence of eye disease other than glaucoma, such as cataract, vascular occlusion or hemorrhage likely to interfere with visual field tests; 4) less than a 7-D refractive error; and 5) HRT images obtained as the mean of three separate scans with an SD less than 50 μm. Patients with a mean deviation (MD) less than −10 dB were excluded.

The results of HRT scans and/or other types of ONH examination were excluded as criteria in making the classification of glaucoma, to make the prediction of visual field test results on the basis of ONH morphology more objective. However, abnormal ONH appearance was often a reason for the initial referral of the patient to the clinic. IOP was not used as a criterion for exclusion or inclusion, because it can be normal in glaucoma (often as a result of ongoing treatment). In cases in which both eyes satisfied the criteria for inclusion, the eye showing the lesser degree of visual field damage was chosen. If no other criteria applied, eyes were chosen to equalize the number of left and right eyes in the sample.

Table 3. Subject Demographics

	Normal	Glaucoma
Total number	100	100
Age range (mean ± SD)	25–87 y (53 ± 14 y)	27–81 y (61 ± 13 y)
Gender (M/F)	43/57	51/49
Eye (R/L)	48/52	51/49
Race		
White	94	87
Asian	6	12
Black	0	1
Mean deviation (dB)	0.3 ± 1.6	−4.9 ± 2.7
SD of HRT scans (mm)	0.024 ± 0.008 mm	0.029 ± 0.010 mm
Refraction (D)	−0.48 ± 2.1 D	−0.61 ± 2.4 D

Data are number of subjects, unless otherwise marked.

Table 3 lists the patient and normal subject demographics for the two groups. The average MD in the glaucoma group was −4.9 dB, which is similar to that in the patient groups studied by Mikelberg et al. (−5.5 dB),[21] Wollstein et al. (−3.6 dB),[32] and Brigatti et al. (−4.5dB).[23]

The study followed the tenets of the Declaration of Helsinki, the subjects provided informed consent, and the project was approved by the Clinical Research Ethics Board of the University of British Columbia.

Classification

The method used to classify images as normal or glaucomatous is based on discriminant function analysis (DFA), and is described in detail in the Appendix. Seven of the parameters were used in the classification. They included the horizontal and vertical components of image curvature, cup radius, maximum cup depth, the temporal cup gradient measure, the fit of the parabolic function, and the fit in the central region. These parameters passed the one-sample Kolmogorov–Smirnov test for normality in both the normal and glaucoma groups. They were used to calculate, for each image, the probability that it came from the glaucoma group, denoted $P(G)$. Cases were classified as normal if $P(G) < 0.5$ and as glaucoma if $P(G) \geq 0.5$.

Comparison with Standard HRT Parameters

To compare the method with the accuracy of classification obtained using the standard HRT parameters, all the images were outlined using standard software provided with the HRT (version 2.01), and the resultant 14 global shape parameters were entered into a spreadsheet. The DFA performed by Mikelberg et al.[21] was then repeated for comparison with the present method. This was accomplished in two ways: by analyzing the original DFA formula values, as incorporated in the HRT software, and by running a separate analysis to generate a new DFA formula with SPSS (ver. 7.5; SPSS, Chicago, IL).

Results

Robustness of the Numerical Procedures

Nonlinear least-squares optimization of parameters is dependent on good initial estimates and may fail to produce mean-

TABLE 4. Model Parameters

Symbol	Description	Normal	Glaucoma	d'	r (age)
c	Horizontal image curvature	0.193 ± 0.091	0.051 ± 0.053	−1.97*	0.16
d	Vertical image curvature	0.045 ± 0.075	−0.049 ± 0.046	−1.55*	0.23†
a	Nasotemporal slant (mm/mm)	−0.093 ± 0.130	0.018 ± 0.097	0.98*	0.002
f	Fit of the model to the image (mm)	0.076 ± 0.021	0.092 ± 0.024	0.71*	−0.061
r_0	Cup radius (mm)	0.444 ± 0.142	0.547 ± 0.154	0.70*	−0.064
z_m	Cup depth (mm)	0.621 ± 0.249	0.739 ± 0.222	0.50*	−0.372*
b	Vertical slant (mm/mm)	0.008 ± 0.072	−0.01 ± 0.068	−0.26	0.005
s	Cup slope	0.116 ± 0.059	0.107 ± 0.055	−0.16	−0.163
z_0	Vertical offset (mm)	0.982 ± 0.312	1.028 ± 0.339	0.14	0.209†

* $P < 0.001$.
† $P < 0.01$.

ingful results if the initial estimates are poor[34] or if the model itself is not a good one. For each image, we therefore checked that the fitting procedure had produced what seemed likely to be correct values, particularly with respect to cup position (x_0 and y_0) and cup depth (z_m). This was achieved by visually examining the model and real images to check that they seemed similar, that cup position was close to the position of the real cup, and that parameter values were within expected ranges. Acceptable fits and parameter values were judged to have been obtained in 198 of the 200 images. The two images in which the automated procedure failed were from normal eyes in which a cup was barely detectable. In one of these images an acceptable fit was obtained by manually choosing initial parameters. In the other image, a fit was obtained by constraining $z_m = 0$ and setting x_0 and y_0 equal to the estimated center of the cup, close to the middle of the image. Radius and slope values were set equal to the means of the rest of the normal group for the purposes of calculating the other derived parameters in this image.

Figure 2 shows examples of 10 real images and the corresponding model images chosen at random: five from the normal group and five from the glaucoma group.

Analysis of Parameter Values

Table 4 gives the means ± SD of the model parameters, and Table 5 gives corresponding values for the morphologic indices derived from them, for both the normal and glaucoma groups. The tables also show a statistical measure of the differences between the groups (d') which is the difference between the means of the two groups divided by the average of the SDs. The rows in each table are in decreasing order of d'—i.e., decreasing statistical difference between the measures in the two groups. A one-way analysis of variance (ANOVA) showed that in almost all cases the differences were statistically significant ($P < 0.001$). The final column, on the right side of each table, shows the Pearson correlation coefficient (r) between each parameter and age, measured in the normal group. Figure 4 shows one-dimensional model profiles, taken along the x-axis at $y = y_0$, calculated using the averages of the parameters for the normal (solid line) and glaucoma (dashed line) groups.

The model generally provided a good description of the images, especially in the normal group, in which the value of the fit (the root mean square of the difference between the model and real image) averaged 0.076 mm. This is also evident from the visual comparisons (Fig. 2). That the fit in the glaucoma group, which averaged 0.092 mm, was less good suggests a correlation between glaucoma and irregularity in cup shape and means that the value of the fit can be used to discriminate between normal and glaucomatous images.

The two model parameters showing the greatest difference between the two groups were horizontal and vertical image curvature, which measure the curvature of the image, minus the model cup, in the x (horizontal or nasotemporal) and y (vertical or superior–inferior) directions, respectively. In most of the normal images both curvature values were positive. It seems likely that these values reflect rim volume (i.e., the increase in thickness of the retinal nerve fiber layer as the axons converge toward the center of the disc). Our measurements show that curvature is greater in the horizontal axis than it is in the vertical axis and that this horizontal component was substantially reduced in the glaucoma group. In normal subjects the vertical component of curvature was smaller than the horizontal component, and in a minority of normal subjects, its value was negative. This component was also substantially reduced in the glaucoma group, where the mean value was negative.

TABLE 5. Secondary Morphologic Indices

Symbol	Description	Normal	Glaucoma	d'	r (age)
g_r^T	Cup gradient measure temporal	8.39 ± 0.58	9.06 ± 0.47	1.28*	−0.15
g_r	Cup gradient measure	9.35 ± 0.49	9.90 ± 0.40	1.23*	−0.18
g_r^N	Cup gradient measure nasal	8.85 ± 0.48	9.31 ± 0.38	1.06*	−0.17
f_R	Fit in central region (mm)	0.107 ± 0.038	0.147 ± 0.048	0.93*	−0.20†
f_P	Fit of parabolic function (mm)	0.137 ± 0.047	0.178 ± 0.054	0.81*	−0.29†
z_{500}	Maximum cup depth (mm)	0.702 ± 0.247	0.878 ± 0.250	0.71*	−0.31†

* $P < 0.001$.
† $P < 0.01$.

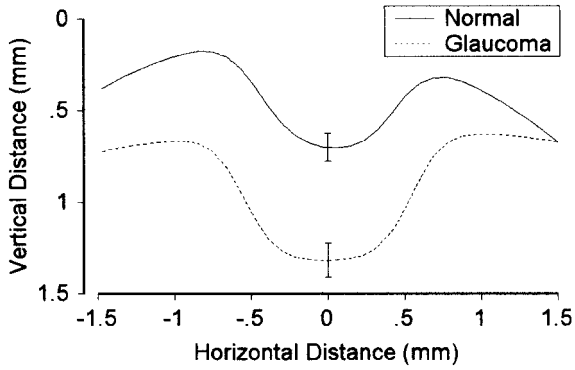

FIGURE 4. One-dimensional profiles through the model function at $y = y_0$—that is, along the horizontal axis through the center of the model cup. The parameters used to calculate the profiles are the means from the normal and glaucoma populations. The vertical offset between the two profiles is arbitrary and was chosen so that they did not cross. Error bars show the root mean square difference between each image and its corresponding model, averaged across all the images in each group.

Vertical slant did not differ significantly from zero and did not differ significantly between the two groups. Nasotemporal slant averaged −0.093 mm/mm (−5.3°) in normal subjects, and this was significantly different from zero. In our sample of glaucoma subjects this difference disappeared and, on average, slant values did not differ significantly from zero. One interpretation of this difference is that there is a relatively greater loss of axons entering the temporal side of the disc in glaucoma. Other interpretations will be discussed later.

The measures of cup depth (z_m) and radius (r_0) were both increased, as would be expected, in the glaucoma group, although these differences were not as large as those for curvature and nasotemporal slant. The model's measure of the slope of the cup walls (s, which is inversely related to the steepness) was slightly smaller in the glaucoma group, reflecting an increased steepness of the walls of the cup. However the difference was not statistically significant, and in the following section we show that other measures of wall steepness were more significantly affected by glaucoma.

The values of f_p, the goodness of fit to the image of a model without a cup (i.e., a parabolic surface), were also analyzed. The fits were relatively poor in most cases, with the exception of images (almost always from normal subjects) in which the cup was poorly defined. Values of f_p differed significantly between the two groups and were found to increase the accuracy with which images could be classified. The value of f_p was lowest in normal eyes in which a cup was barely detectable or absent. Because it can be calculated without the need for initial guesses of parameter values, a low value of f_p (e.g., <0.075 mm) can be used to identify images without a cup. Although it was not done here, these images can safely be assumed to be normal and can be excluded from further processing.

Although the model describes normal and abnormal discs relatively well and many of the model parameters differed significantly between groups, it was clear that it failed to describe some significant features of glaucomatous discs, in particular the notably increased steepness of the cup walls. We therefore calculated additional morphologic indices from the images, using the model parameters as a framework for the calculations. The center of the model cup, its radius, and slope were used to define a central circular region of the image, denoted R, which just enclosed the cup and its walls (Fig. 3). Visual checks were made, and it was found that in no case did any part of a cup appear to fall outside this region, nor, in most cases, did the region greatly exceed the cup in size. Three measures of steepness were calculated, one for the whole region (g_r) one for the nasal half of the region (g_r^N), and one for the temporal half (g_r^T). Table 5 shows that, in the normal group, the nasal gradient measure was larger than the temporal gradient measure, possibly because of the greater number of blood vessels on the nasal side. However the temporal measure differed more between the normal and the glaucoma groups.

An additional measure of goodness of fit was made: f_R measured the fit of the model function within R. This value was significantly larger, on average, in images from the glaucoma group (Table 5).

The measure of cup depth (z_m) may be relatively insensitive to small local excavations in the bottom of the cup, which may be indicative of glaucoma. Therefore, we took as a depth measure the average of the 500 deepest values measured within region R (which typically contained 2500–3000 pixels). This measure (z_{500}) with $d' = 0.71$, differed more between the two groups than did z_m, for which $d' = 0.50$.

Effects of Age

The effect of age on the parameters was examined in the normal group by calculating the Pearson correlation between each parameter and age. The results are shown in Tables 4 and 5. Age had a significant effect on cup depth, where the correlation ($r = -0.372$, slope = −0.0068 mm/y) indicates a decrease in depth with increasing age. A smaller positive correlation between age and the horizontal and vertical components of curvature was also found. In almost every case, the effects of age, although small in magnitude, were in the direction opposite those of glaucoma. This should result in an increasing dissimilarity between normal and glaucomatous discs with increasing age and, at least in theory, should make the detection of glaucoma easier in older subjects.

Classification

Figure 5 shows the distribution of P(G) values (i.e., the probability that an image comes from the glaucoma population) for the two populations. It shows that most cases were correctly classified with high confidence levels (i.e., $P > 0.9$ or < 0.1). Table 6 shows the distribution of probability values and of classification mistakes. The overall classification accuracy was 89% (specificity, 89%; sensitivity, 88%). As might be expected, the accuracy varied with the confidence level of the classification. When the confidence level was low (i.e., for P between 0.4 and 0.6; leftmost column in Table 5), the accuracy was 67% (4/6 cases). For high confidence levels (i.e., $P > 0.9$ or < 0.1; rightmost column in Table 6), the overall accuracy was higher at 96% (118/123 cases). Eighty-seven percent of cases were classified with $P > 0.7$ or < 0.3. Within this group, the overall accuracy was 92%.

We performed a jackknife (leave-one-out) cross-validation procedure in which data from the case classified were not used in the calculation of group means and covariance values. This gives a better estimate of the method's ability to generalize to new cases—that is, cases not used to derive the classification

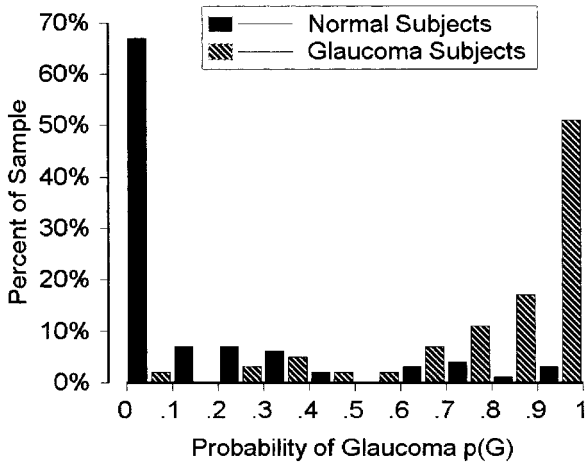

FIGURE 5. The distribution of the values of *P*(G) in the normal and glaucoma groups.

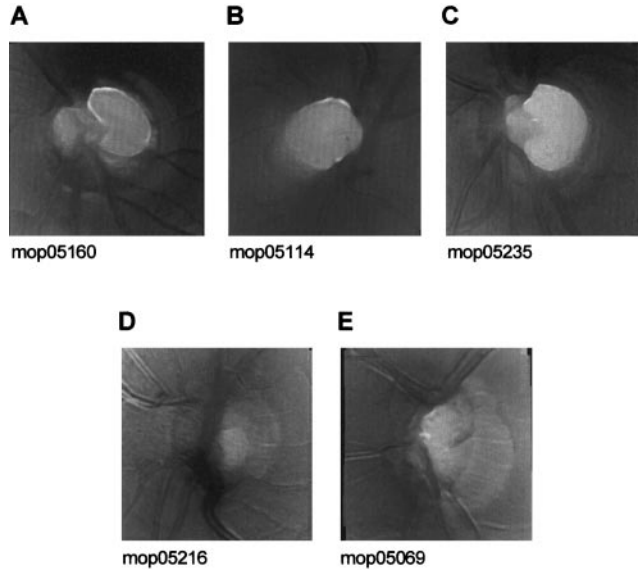

FIGURE 6. (**A**, **B**, and **C**) False-positive images from the normal group, classified with $P(G) > 0.9$; (**D, E**) False-negative images from the glaucoma group, classified with $P(G) < 0.1$. Brightnesses and contrasts of individual images have been adjusted for clarity.

function. With this procedure, six additional wrong classifications were identified, and the overall accuracy of the method was reduced to 86%. However, of the new mistakes, three had *P* values that fell in the 0.4 to 0.6 range, two had *P* values in the 0.2 to 0.4 range, and only one had $P > 0.7$. Overall, the accuracy of the confidently classified cases was reduced to 88%.

We examined, retrospectively, those images that had been confidently misclassified by the procedure (i.e., those with $P < 0.1$ or $P > 0.9$), as well as the corresponding visual fields. This included three normal and two glaucoma subjects (Fig. 6). The clinical interpretation of the normal cases was that two of them had large discs, suggesting that the large cups were a consequence of large disc size. The interpretation of the appearance of the third disc was that it was suspicious, although the visual field was normal. In both confident false-negative glaucoma cases, the clinical interpretation was of normal disc appearance despite glaucomatous visual fields. The observation that the false-positive cases had large discs suggested that this might account for some of the other false positive results. Analysis of disc area (i.e., the HRT parameter *ag*) showed that disc area in the 11 false-positive cases was 3.07 ± 0.570 mm^2, and this was significantly larger ($P < 0.001$) than the area in the correctly classified normal subjects, which was 2.335 ± 0.578 mm^2.

Comparison with Standard HRT Parameters and Visual Field Indices

We compared the accuracy of the present method with that which could be obtained from the standard set of shape parameters calculated by the HRT operating software. The DFA formula of Mikelberg et al.[21] which is incorporated into the software (ver. 2.01) gave a sensitivity of 49% and a specificity of 98%. Adjusting the classification threshold to give more equal values yielded a sensitivity of 77% and a specificity of 77%. A more fair comparison is to subject the data to a new DFA. Thirteen of the parameters (*ag* [disc area] was excluded because it is derived from *mr* [mean radius]) were entered into a forward-stepping DFA, using an F-to-enter of 4.0 and an F-to-remove of 3.0. The overall accuracy was 84%, and the cross-validated accuracy was 83.5%. The six parameters selected by the analysis were *abr* (area below reference), *mhc* (mean height of contour), *mr* (mean radius), *var* (volume

TABLE 6. Classification Statistics

	According to Range of *P*				
	$0.4 < P < 0.6$	$0.6 < P < 0.7$ $0.3 < P < 0.4$	$0.7 < P < 0.8$ $0.2 < P < 0.3$	$0.8 < P < 0.9$ $0.1 < P < 0.2$	$0.9 < P < 1.0$ $0.0 < P < 0.1$
Number of cases	6	21	25	25	123
Number of mistakes	2	8	7	1	5
Correct (%)	67	62	72	96	96
	According to Cumulative *P*				
	$0.0<P<1.0$	$0.6<P<1.0$ $0.0<P<0.4$	$0.7<P<1.0$ $0.0<P<0.3$	$0.8<P<1.0$ $0.0<P<0.2$	$0.9<P<1.0$ $0.0<P<0.1$
Total cases (%)	100	97	87	74	62
Correct (%)	89	89	92	96	96

TABLE 7. Correlations with Standard HRT Parameters, Visual Field MD, and ROC Areas

	ag	*abr*	*mhc*	*hvc*	*mdg*	*csm*	*var*	*vbr*	r_0	z_{500}	*c*	*d*	g_r^T	f_R	f_P
abr	0.68														
mhc	0.22	0.34													
hvc	−0.04	−0.36	−0.22												
mdg	0.48	0.70	0.32	0.03											
csm	0.21	0.41	0.12	−0.21	0.18										
var	0.12	−0.52	−0.29	0.73	−0.27	−0.25									
vbr	0.51	0.82	0.24	−0.19	0.66	0.31	−0.38								
r_0	0.52	0.58	0.20	−0.12	0.31	0.37	−0.10	0.43							
z_{500}	0.52	0.66	0.24	0.13	0.94	0.19	−0.13	0.64	0.32						
c	−0.34	−0.64	−0.53	0.54	−0.47	−0.27	0.62	−0.41	−0.44	−0.32					
d	−0.30	−0.60	−0.53	0.38	−0.58	−0.22	0.57	−0.42	−0.28	−0.39	0.84				
g_r^T	0.51	0.78	0.31	−0.16	0.75	0.40	−0.40	0.61	0.51	0.78	−0.55	−0.55			
f_R	0.37	0.60	0.21	0.09	0.85	0.18	−0.20	0.59	0.33	0.82	−0.34	−0.44	0.68		
f_P	0.54	0.72	0.25	0.07	0.94	0.28	−0.18	0.72	0.44	0.94	−0.38	−0.45	0.77	0.83	
MD	−0.09	−0.43	−0.41	0.32	−0.27	−0.23	0.48	−0.24	−0.29	−0.20	0.55	0.48	−0.46	−0.27	−0.26
ROC area	0.60	0.85	0.91	0.74	0.73	0.77	0.85	0.84	0.69	0.69	0.93	0.86	0.82	0.76	0.73

above reference), *vas* (volume above surface), and *vbr* (volume below reference).

We calculated the correlations, across both normal and glaucoma groups, between the model parameters and the standard HRT parameters, and with the visual field MD. Table 7 shows the values for some selected HRT and model parameters. As might be expected, the HRT measure of disc area (*ag*) did not correlate strongly with any of the model parameters, because the model does not provide an explicit estimate of disc area. The highest correlation ($r = 0.54$) was with the fit of the parabolic function (f_P). However, there was also a strong correlation ($r = 0.52$) with the model's measure of cup radius (r_0) which can be explained because cup area and disc area are known to be strongly correlated in normal discs.[37–39] There was a strong correlation ($r = 0.94$) between the model's measure of cup depth (z_{500}) and the corresponding HRT measure (*mdg*). The fit of the parabolic surface (f_P) also correlated strongly ($r = 0.94$) with *mdg*, because its value is small when a cup is absent, and therefore its value largely reflects cup depth. Both z_{500} and r_0 showed weak correlations with HRT parameters *hvc* and *var*, on which the shape of the cup would be expected to have little effect.

The HRT parameter *csm* (cup shape measure) has been shown by several studies to give good discrimination between groups, although its structural interpretation seems obscure. The model parameter that correlated most strongly with *csm* was temporal gradient measure, g_r^T ($r = 0.40$), however the correlation with cup diameter (r_0) was nearly as strong ($r = 0.37$).

We calculated the correlation between each parameter and the visual field MD. These values are also given in Table 7 (second row from the bottom). The parameter showing the highest correlation with MD was horizontal image curvature (*c*; $r = 0.55$). The HRT indices showing the strongest correlations were *abr* ($r = -0.43$), *var* ($r = 0.48$), and *mhc* ($r = -0.41$).

For comparison with previous studies, we subjected all the HRT and model parameters to a receiver operating characteristic (ROC) analysis, using the methods described in Iester et al.[27] Some of these values are given in the bottom row of Table 7. The parameter showing the highest area under the curve (a measure of discrimination between two groups) was the horizontal curvature measure (*c*, area = 0.93). The best HRT measure was *mhc* (area = 0.91); *csm* did less well, with an area = 0.77.

DISCUSSION

Our results show that it is possible to analyze and classify images of the ONH with a degree of reliability that is comparable to that of existing methods, without the need for prior manual outlining of the optic disc. The method that we propose takes only a few seconds to perform on a computer and requires no user intervention at any stage in the numerical analysis. As such, it is reliable and repeatable and is not dependent on subjective estimates of the position of the borders of the optic disc. These estimates may vary within and between operators and complicate the comparison of studies conducted in different centers. Although the details of the method itself are somewhat technical, the structural interpretation of the parameters extracted by it is less complex. The measures identified by the model as particularly useful in indicating glaucomatous damage to the disc include the horizontal and vertical components of image curvature (i.e., the amount by which the nerve fiber layer surrounding the cup bulges upward into the vitreous) and the steepness of the cup walls. Cup size and other measures of surface irregularity in the region of the cup are also informative. Before discussing the interpretation of these shape changes in more detail, we will consider first the limitations of the mathematical modeling method we have used and, secondly, the limitations imposed by the study design.

Limitations of the Method

The particular mathematical model that we have chosen is unlikely to be the only one that could be used. Its two main components, a circularly symmetric cup placed on a background with parabolic curvature, are both arguably unrealistic choices that could be improved, given that the goal is to find a model that can accurately reproduce the range of ONH shapes encountered in normal subjects. First, real cups are often not circularly symmetric. Although modeling an asymmetric cup would require introduction of additional shape parameters, some of them might turn out to be informative. Second, the parabolic curvature of the background (i.e., the rim and disc

margins) is unrealistic, because it leads to depth values that increase as the square of the distance from the center of the cup, which obviously does not happen in reality. The function probably only works because the images, which are 10° × 10° in size, do not extend outside the region in which the retinal nerve fiber layer is becoming increasingly thick as fibers converge toward the optic nerve. This suggests that the method would work less well on larger (e.g., 15° × 15°) images.

Finding the best-fitting model parameters for each image requires the application of iterative nonlinear least-squares optimization, which is not guaranteed to work in all cases.[34] Although we chose a method (Levenburg–Marquardt) that is believed to be one of the most efficient, it, as in all similar procedures, requires a good initial estimate of the parameters that are to be adjusted if it is to work properly. It is not hard to make these estimates, however, and in our sample of images the method almost always (198 of 200 cases) converged on an acceptable solution. The two images in which the method failed were both from normal eyes in which a cup was barely perceptible. Such cases can be detected either by visual inspection or by first fitting a function without a cup—that is, by calculating f_p. If this value is low (e.g., <0.075 mm) the image can, on the basis of the present results, be safely classified as normal (only one of the glaucoma images had a value of $f_p <$ 0.10 mm, and this image was classified as normal by the subsequent analysis in any case).

The model does not provide any estimate that can be directly related to disc area. Some of its parameters correlate with the HRT measure of area (Table 7), but these correlations seem likely to be indirect. Disc area itself is not affected in glaucoma[29,40]; however, cup area and disc area are strongly correlated in normal eyes[37–39] and the model's measure of cup size would probably be more useful if it could be accompanied by an estimate of disc area and be re-expressed as a ratio. This is supported by the observation that disc area tends to be larger in the misclassified normal subjects. It remains to be seen whether such an estimate can be provided by an automated method (perhaps by analyzing the reflectivity image, using the determination of cup size and position by the present method as a guide).

The classification method we used is based on the assumption that the data values in each group are distributed according to a multivariate normal distribution. It is related to the methods used in DFA,[41] but we have also used the advantage of providing an estimate of the probability that an image belongs to one of the two groups. We have not so far tested other classification methods that could be used (e.g., back-propagation neural nets[23,42]) that might perform better than the present method.

Limitations of the Study

Validation of the accuracy of the classification method depends critically on the selection of subjects for the control normal and glaucoma groups. Ideally, the normal subject group should be an unbiased sample of the glaucoma-free population, and the subjects in the glaucoma group should also be an unbiased sample of the glaucoma population and have early visual field damage. We cannot guarantee that either of these conditions has been met in the present study. Although we took pains to exclude the presence of glaucoma in our normal group by means of visual field testing and IOP measurements, an unusually high proportion of the volunteers (33/100) reported a family history of glaucoma. Often, this was a reason for volunteering. This raises the possibility that some of them may either have had early glaucoma or that they may have had a higher proportion of congenital disc abnormalities predisposing them to glaucoma than would be found in the normal population. This would tend to decrease rather than increase the classification accuracy of the study. As it turned out, the proportion of false-positive cases within the family history group (3/33) was slightly less than it was in the rest of the group (8/67). It can be argued that it is not necessarily bad that screening methods be tested with a population at risk for glaucoma (e.g., elderly people with a family history), provided glaucoma has been excluded as best it can without ONH examination, because it is persons in this population who are most likely to seek screening.

It is possible that optic disc morphology in our glaucoma sample was more abnormal than it is in the glaucoma population as a whole, because an abnormal disc morphology may have been one of the reasons for initial referral to the glaucoma clinic. (The implication is that glaucoma in combination with a relatively normal disc appearance exists and often goes undiagnosed.) This could artifactually increase the accuracy of our classification. There seems little we can do about this source of bias that will persist even though we were careful to apply only visual field criteria in selecting patients for inclusion or exclusion.

We attempted to match the samples for age, excluding normal subjects less than 25 years of age, to make the distributions match more closely. The normal group was nevertheless approximately 8 years younger (approximately 0.6 SDs) than the glaucoma group. Age was found to have a significant effect on some of the parameters used in the classification. This could bias the results, because age alone would then produce a difference in the parameter values between the two groups. However, when we examined the effect of age on normal discs we found that in almost every case the effects were in the direction opposite those found in the glaucoma group. In other words, in normal subjects, discs become less glaucomatous in appearance as they age, and discrimination between the two groups should be easiest in older subjects. The reasons for this do not seem clear. The difference in age between the two groups is likely, however, at least in theory, to have made classification more difficult.

Interpretation of the Shape Changes

The two parameters showing the largest difference (in statistical terms) between the two groups were horizontal and vertical image curvature (parameters *c* and *d*, Table 4). Positive values of these indices mean that the neuroretinal rim region around the cup is convexly curved, causing it to bulge upward into the vitreous. The degree of curvature seems likely to be determined at least in part by geometric factors. As axons converge toward the center of the disc the nerve fiber layer will become, of necessity, increasingly thick. Only as axons leave the disc through the optic nerve can the layer start to decrease in height. On geometric grounds curvature would be predicted to be proportional to the total number of ganglion cell axons and inversely proportional to the diameter of the nerve. Our results bear this out, inasmuch as the curvature values were greatly reduced in the glaucoma patients (suggesting a substantial loss of axons), whereas a negative correlation between *c* and *d* and disc area (*ag*; $r = -0.34$ and -0.30,

respectively) was also observed (Table 7). These considerations suggest that it may be of interest to perform a more exact mathematical analysis of the way in which geometric factors determine ONH surface topography.

The horizontal component of curvature was greater than the vertical component in normal ONH images, and this implies that the rim region tends to be elongated vertically. In images in which the horizontal component is positive and the vertical component negative (some normal images, and most glaucoma ones) the rim region is saddle-shaped and can be pictured as two parallel ridges on the nasal and temporal sides of the cup. That both values were decreased by relatively large amounts in the glaucoma group (Table 4) suggests that axons had been lost from many different regions of the retina. The observation of negative curvature values along the vertical axis in the glaucoma group suggests a relatively large loss of axons in the superior and inferior regions of the ONH; however, that the horizontal component changes more in absolute terms than the vertical component seems to suggest the opposite—that is, a greater number of axons entering the nerve on the nasal and temporal sides has been lost. It is possible that this loss is less apparent because there were more axons in these regions to begin with.

Previous morphologic studies of the ONH in glaucoma do not appear to have attempted to quantify the steepness of the cup walls. The measure we devised (g_r; Fig. 3) showed a statistically large difference between groups (Table 5). We defined g_r in terms of the component of the gradient measured in a direction radial to the center of the cup (Appendix), because our initial measurements showed that although similar noncomponent measures were significantly greater in the glaucoma group, measures based on the radial component differed more. It is in this direction that the pressure gradient across the cup is steepest,[43] and our findings are consistent with the expectation that pressure is one of the sources of damage to the nerve. However, an increase in the steepness of the cup walls may be the result of glaucoma, as well as a cause of it. For example, focal loss of ganglion cells originating in one region of the retina would cause the disappearance of a bundle of axons originating from that region and may be manifest as a notch in the rim, which in turn could lead to a steepening of the cup wall at that location. Alternatively, congenitally steep regions of the cup could predispose toward glaucomatous damage, given that the steeper the gradient the smaller the radius of curvature of the axons as they enter the nerve. Such sharp bends may be especially vulnerable to pressure-induced damage. It is possible that both mechanisms are at work, leading to a vicious cycle of damage.

We further divided the gradient measure into nasal and temporal components. Our results showed that in normal images the nasal component was greater on average than the temporal component (Table 5). However, the temporal component was more severely affected by glaucoma. It is possible that the nasal component is more influenced by blood vessels, which are more numerous on the nasal side and which would probably not be affected by glaucoma. The temporal component would then be a better measure, whether or not it was in fact more severely affected by glaucoma. We have not so far performed a more detailed sector-based examination of gradient measures, but this could be worthwhile.

Other measures showed smaller differences between the two groups, although they were also useful in classification. The measures r_0 and z_m were both increased in the glaucoma group, although, as has been observed previously,[44] they are relatively poor indicators on their own. When analyzing cup depth measures, we found that z_m differed less between the groups than did the alternative measure (z_{500}), which is an average of the 500 most extreme depth values within the cup region. This measure would be expected to be sensitive to the presence of localized excavations in the cup—namely, the pitting that is observed in glaucoma.[45]

The interpretation of the goodness-of-fit parameter f_R (root mean square difference between the model and the image in the cup region) is less clear. Errors in the determination of topography by the HRT are likely to contribute only a small amount to the values, because these errors (the pixel variability we observed on repeat imaging) average approximately ± 0.025 mm, as has been shown by others[36] and in the present results, and the values of f_R were almost always several times greater than this (Table 5). Although it seems clear that glaucomatous ONH images show shape variations that are not captured by the model and that are not present in normal images, we have not identified what these differences are. There are many possibilities, including notches and variations in the height of the rim that are not captured by the curvature measures, asymmetric cup shapes, excavations in the walls and floor of the cup, and irregularities caused by blood vessels. Further characterization of these variations should be possible, and may be more informative than the goodness-of-fit measures.

One other measure, slant in the nasotemporal axis, showed a significant difference between the two groups (Table 4). Normal images tended to be slanted, by approximately 6° (-0.098 mm/mm), on average, in such a way that the temporal (foveal) side is higher than the nasal side. This slant was usually absent in the glaucoma group. The slant in normal subjects may be explained by the fact that more axons enter the disc on the temporal than on the nasal side, causing an overall negative slant of the surface in the nasal direction. This effect would be offset by the greater number of blood vessels on the nasal side of the disc that may increase the overall measure of height on that side.[46] Although the decrease in slant observed in the glaucoma group suggests a relatively greater loss of fibers entering the nerve on the temporal side (i.e., on the side closest to the fovea), the difference should be interpreted cautiously. Slant in the image can be introduced by changes in the angle of the ophthalmoscope relative to the optical axis of the eye[18] and this angle is chosen by the operator. The images from the normal group were all obtained by one operator, who was not one of those who obtained the images in the glaucoma group. Systematic differences in the operator settings, although small, may have caused artifactual differences between the groups. Against this, it could be argued that when imaging is performed through undilated pupils (as was the case in this study) there is very little scope for changing the viewing angle of the ophthalmoscope. In practice, once the subject is fixating, the angle is determined almost entirely by the requirement that the disc be close to the center of the image.

This study, like others before it,[21,23,27,28,31,32,42,44] has shown that ONH profiles of patients with glaucoma can be distinguished from those of normal subjects with a high degree of reliability. Although it has identified several new shape parameters that are substantially altered by glaucoma, it has not shown what the pattern of change is in the very earliest stages

of glaucoma, before visual field changes have occurred. It is plausible that those factors that are most changed by glaucoma are the ones that change earliest, but this is not necessarily the case. The early pattern of change may be different and may be harder to detect reliably. Loss of surface curvature, steepening of the cup walls, greater surface irregularity, and pitting, widening, and deepening of the cup are all different types of change that may not all occur at the same time or to the same degree in different patients. Automated mathematical analysis of ONH shape, because it does not depend on subjective estimates of disc boundaries, offers a promising method for identifying and reliably quantifying these different changes in long-term studies in individual patients.

Acknowledgments

The authors thank the subjects who generously volunteered for this study; Phil Hetherington and Tim Blanche for useful suggestions; and Gerhard Zinser of Heidelberg Engineering for his comments on the manuscript and for kindly supplying the programs HRTCOMP and DBSCALES.

References

1. Quigley HA. Open angle glaucoma. *N Engl J Med.* 1993;328:1097–1106.
2. Sommer A. Glaucoma: facts and fancies. *Eye.* 1996;10:295–301.
3. Pederson JE, Anderson DR. The mode of progressive disc cupping in ocular hypertension and glaucoma. *Arch Ophthalmol.* 1980;98:490–495.
4. Quigley HA, Addicks EM, Green WR. Optic nerve damage in human glaucoma, III: quantitative correlation of nerve fibre loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. *Arch Ophthalmol.* 1982;100:135–146.
5. Sommer A, Pollack I, Maumenee AE. Optic disc parameters and onset of glaucomatous field loss. *Arch Ophthalmol.* 1979;97:1444–1448.
6. Sommer A, Katz J, Quigley HA, et al. Clinically detectable nerve fibre atrophy precedes the onset of glaucomatous field loss. *Arch Ophthalmol.* 1991;109:77–83.
7. Weinreb RN, Dreher AW, Bille J. Quantitative assessment of the optic nerve head with the laser tomographic scanner. *Int Ophthalmol.* 1989;13:25–27.
8. Kruse FE, Burk RO, Völcker HE, Zinser G, Harbarth U. Reproducibility of topographic measurements of the optic nerve head with laser tomographic scanning. *Ophthalmology.* 1989;96:1320–1324.
9. Dreher AW, Tso PC, Weinreb RN. Reproducibility of topographic measurements of the normal and glaucomatous optic nerve head with the laser tomographic scanner. *Am J Ophthalmol.* 1991;111:221–229.
10. Cioffi GA, Robin AL, Eastman RD, et al. Confocal laser scanning ophthalmoscope: reproducibility of optic nerve head topographic measurements with the confocal scanning laser ophthalmoscope. *Ophthalmology.* 1993;100:57–62.
11. Mikelberg FS, Wijsman K, Schulzer M. Reproducibility of topographic parameters obtained with the Heidelberg Retina Tomograph. *J Glaucoma.* 1993;2:101–103.
12. Lusky M, Bosem ME, Weinreb RN. Reproducibility of optic nerve head topography measurements in eyes with undilated pupils. *J Glaucoma.* 1993;2:104–109.
13. Rohrschneider K, Burk RO, Völcker HE. Reproducibility of topographic data acquisition in normal and glaucomatous optic nerve heads with the laser tomographic scanner. *Graefes Arch Clin Exp Ophthalmol.* 1993;231:457–464.
14. Rohrschneider K, Burk RO, Kruse FE, Völcker HE. Reproducibility of the optic nerve head topography with a new laser tomographic scanning device. *Ophthalmology.* 1994;101:1044–1049.
15. Bartz–Schmidt KU, Weber J, Heimann K. Validity of two dimensional data obtained with the Heidelberg retina tomograph as verified by direct measurements in normal optic nerve heads. *Ger J Ophthalmol.* 1994;3:400–405.
16. Chauhan BC, LeBlanc RP, McCormick TA, Rogers JB. Test–retest variability of topographic measurements with confocal scanning laser tomography in patients with glaucoma and control subjects. *Am J Ophthalmol.* 1994;118:9–15.
17. Janknecht P, Funk J. Optic nerve head analyzer and Heidelberg retina tomograph: accuracy and reproducibility of topographic measurements in a model eye and in volunteers. *Br J Ophthalmol.* 1994;78:760–768.
18. Orgül S, Cioffi GA, Bacon DR, Van Buskirk M. Sources of variability of topometric data with a scanning laser ophthalmoscope. *Arch Ophthalmol.* 1996;114:161–164.
19. Burk ROW, Rohrschneider K, Noak H, Völcker HE. Volumetrische Papillenanalyse mit Hilfe der Laser-Scanning Tomographie: Parameterdefinition und Vergleich von Glaukom und Kontrolpapillen. *Klin Monatsbl Augenheilkd.* 1991;198:522–529.
20. Brigatti L, Caprioli J. Correlation of visual field with scanning confocal laser optic disc measurements in glaucoma. *Arch Ophthalmol.* 1995;113:1191–1194.
21. Mikelberg FS, Parfitt CM, Swindale NV, Graham SL, Drance SM. Ability of the Heidelberg Retina Tomograph to detect early glaucomatous visual field loss. *J Glaucoma.* 1995;4:242–247.
22. Weinreb RN, Shakiba S, Sample PA, et al. Association between quantitative nerve fibre layer measurement and visual field loss in glaucoma. *Am J Ophthalmol.* 1995;120:732–738.
23. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol.* 1996;121:511–521.
24. Uchida H, Brigatti L, Caprioli J. Detection of structural damage from glaucoma with confocal laser image analysis. *Invest Ophthalmol Vis Sci.* 1996;37:2393–2401.
25. Hatch WV, Flanagan JG, Etchells EE, Williams-Lyn DE, Trope GE. Laser scanning tomography of the optic nerve head in ocular hypertension and glaucoma. *Br J Ophthalmol.* 1997;81:871–876.
26. Iester M, Mikelberg FS, Courtright P, Drance SM. Correlation between the visual field indices and Heidelberg retina tomograph parameters. *J Glaucoma.* 1997a;6:78–82.
27. Iester M, Mikelberg FS, Swindale NV, Drance SM. ROC analysis of Heidelberg retina tomograph optic disc shape measures in glaucoma. *Can J Ophthalmol.* 1997b;32:382–388.
28. Iester M, Mikelberg FS, Drance S. The effect of optic disc size on diagnostic precision with the Heidelberg Retina Tomograph. *Ophthalmology.* 1997c;104:545–548.
29. Iester M, Swindale NV, Mikelberg FS. Sector-based analysis of optic nerve head shape parameters and visual field indices in healthy and glaucomatous eyes. *J Glaucoma.* 1997d;6:371–376.
30. Anton A, Yamagishi N, Zangwill L, Sample PA, Weinreb RN. Mapping structural to functional damage in glaucoma with standard automated perimetry and confocal scanning laser ophthalmoscopy. *Am J Ophthalmol.* 1998;125:436–446.
31. Bathija R, Zangwill L, Berry CC, Sample PA, Weinreb RN. Detection of early glaucomatous structural damage with confocal scanning laser tomography. *J Glaucoma.* 1998;7:121–127.
32. Wollstein G, Garway–Heath DF, Hitchings RA. Identification of early glaucoma cases with the scanning laser ophthalmoscope. *Ophthalmology.* 1998;105:1557–1563.
33. Orgül S, Cioffi GA, van Buskirk EM. Variability of contour line alignment on sequential images with the Heidelberg Retina Tomograph. *Graefes Arch Clin Exp Ophthalmol.* 1997;235:82–86.
34. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes: The Art of Scientific Computing*. 2nd ed. Cambridge, UK: Cambridge University Press; 1994.
35. Bishop CM. *Neural Networks for Pattern Recognition.* Oxford University Press; 1995.
36. Weinreb RN, Lusky M, Bartsch D, Morsman D. Effect of repetitive imaging on topographic measurements of the optic nerve head. *Arch Ophthalmol.* 1993;111:636–638.
37. Teal PK, Morin JD, McCulloch C. Assessment of the normal disc. *Trans Am Ophthalmol Soc.* 1972;70:164–177.
38. Bengtsson B. The variation and covariation of cup and disc diameters. *Acta Ophthalmol.* 1976;54:804–818.

39. Britton RJ, Drance SM, Schulzer M, Douglas GR, Mawson DK. The area of the neuroretinal rim of the optic nerve in normal eyes. *Am J Ophthalmol.* 1987;103:497–504.
40. Jonas JB, Gusek GC, Naumann GOH. Optic disc morphometry in chronic primary open-angle glaucoma. *Graefes Arch Clin Exp Ophthalmol.* 1988;226:531–538.
41. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 2nd ed. New York: Harper Collins; 1989.
42. Parfitt CM, Mikelberg FS, Swindale NV, Green S, Graham SL, Drance SM. The use of artificial neural networks to identify optic nerve head degeneration due to glaucoma [ARVO Abstract]. *Invest Ophthalmol Vis Sci*. 1995;36(4):S628.
43. Fechtner RD, Weinreb RN. Mechanisms of optic nerve damage in primary open angle glaucoma. *Surv Ophthalmol.* 1994;39:23–42.
44. Caprioli J. Discrimination between normal and glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 1992;33:153–159.
45. Radius RL, Maumenee AE, Green WR. Pit-like changes of the optic nerve head in open-angle glaucoma. *Br J Ophthalmol.* 1978;62:389–393.
46. Dichtl A, Jonas JB, Mardin CY. Comparison between tomographic scanning evaluation and photographic measurement of the neuroretinal rim. *Am J Ophthalmol.* 1996;121:494–501.

APPENDIX

The equation used to describe the surface topography of the ONH was:

$$z(x, y) = \frac{z_m}{1 + e^{(r-r_0)/s}} + a(x - x_0) + b(y - y_0) + c(x - x_0)^2 + d(y - y_0)^2 + z_0 \quad (1)$$

where

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (1a)$$

This defines the depth of the surface (z) as a function of position (x, y) on the surface. Movement in the positive x direction is identified with movement in the nasal direction, away from the fovea, and movement in the positive y direction corresponds to movement toward the superior retina. The first term on the righthand side of the equation represents a circularly symmetric cup, centered on position (x_0, y_0), with a depth z_m, a radius r_0, and walls with a slope inversely proportional to s (the smaller the value of s, the steeper the slope). The following four terms describe a surface with variable slant in the x and y directions (parameters a and b, respectively) and with variable curvature, assumed to be parabolic, in the x and y directions (parameters c and d, respectively). The radius (z_0) is a constant offset in the z direction.

For each image, the 10 free parameters of the model were adjusted to give the best fit of the model to the image. Fit (f) was defined as the root mean square of the difference between the image and the model, measured in millimeters

$$f = \left\{ \frac{1}{N} \sum_{i,j}^{all\ pixels} [\gamma I(i, j) - z(\alpha i, \beta j)]^2 \right\}^{1/2} \quad (2)$$

where $I(i,j)$ is the value of the image at pixel (i,j) and N is the total number of pixels in the image. The terms α and β, which are normally equal, scale pixel indices in the i and j directions to millimeters in the x and y directions, respectively, and γ scales the one-byte-per-pixel value in the image (0–255) to millimeters in the depth (z) dimension.

Fitting was performed in two stages: first, we made initial estimates of the parameter values and second, we refined the values to minimize f, using an iterative nonlinear least-squares fitting procedure. The initial estimates were made as follows: 1) we calculated a least-squares fit to the image of just the last five terms of equation 1 (i.e., of the parabolic surface). Because this function is linear in its five parameters, it is possible to calculate the best fitting parameter values explicitly using standard methods (Gauss–Jordan elimination).[34] 2) This function was then subtracted from the raw image, to obtain one in which the cup should be the major feature. The average of the positions and values of the 50 largest (i.e., deepest) pixel values in this image was then calculated to obtain estimates of cup position (x_0,y_0) and depth (z_m), respectively. A region of pixels, equal to one tenth of the image width, along the edges of the image was excluded when searching for deep pixels. 3) The fit of the parabolic surface was then repeated, this time excluding the region of the image likely to contain the cup—namely, a region within a distance of 0.5 mm from the estimated center of the cup. This fit was used to obtain the initial estimates of a, b, c, d, and z_0. The initial estimates of cup radius and slope were fixed at 0.5 mm and 0.1, respectively. After this, the parameter estimates were further refined using the Levenburg–Marquardt optimization technique.[34]

These procedures were applied to 10° × 10° images extracted from databases created by the operating software (version 2.01) provided with the HRT. The program HRTCOMP was used to extract the images, and the program DBSCALES was used to extract the appropriate scaling parameters (α, β, and γ) for each image. Because the edges of the images sometimes contain artifacts, a region typically 10 pixels wide along each edge of the image was excluded from analysis. To decrease processing time, the 256 × 256-pixel images were reduced in size by averaging over blocks of 4 × 4 pixels, to give typically 60 × 60-pixel images. Averaging over smaller blocks, or not averaging at all, made little difference to the estimated parameter values.

Morphologic Indices Derived Using the Model

After the function fits and derivation of parameter values, additional morphologic indices were calculated. The selection and definition of these was guided by their usefulness in discriminating between normal and glaucomatous images. First, a set of pixels, R, which included the cup was defined with a center position (x_0, y_0) and a radius = $r_0 + \log_e(9)s$. Within this region, cup depth is greater than 10% of its value at the center. The calculations were performed on the raw 256 × 256-pixel images. As described, pixels on a defined border along the edges of the image were excluded from R. The following values were then calculated.

First, the goodness of fit was defined

$$f_R = \left\{ \frac{1}{N} \sum_{i,j \subset R} [\gamma I(i, j) - z(\alpha i, \beta j)]^2 \right\}^{1/2} \quad (3)$$

where N is the number of points contained in R. The larger the value of f_R, the worse is the fit of the model to the image in the region of the cup. As described above, this value was found to be significantly larger in eyes with glaucoma.

Second, an index of the steepness of the cup walls was determined. Although parameter s (equation 1) gives a measure of steepness, we found that a more informative measure could be obtained by summing the image gradient values within R. This was performed with two further modifications: Only the radial component of the gradient (i.e., the component measured in the direction pointing toward the center of the cup) was used, and only gradient values with large negative slope values less than $-45°$ were included in the sum. We define the gradient in terms of its x and y components as

$$G_x(i, j) = \frac{\gamma[I(i + 1, j) - I(i, j)]}{\alpha}; \qquad G_y(i, j) = \frac{\gamma[I(i, j + 1) - I(i, j)]}{\beta}. \tag{4}$$

The radial component G_r is given by

$$G_r = \frac{(x - x_0)G_x + (y - y_0)G_y}{|r|}. \tag{5}$$

Because depth is measured as a positive quantity, the steeper the slope of the cup walls, the more negative are the radial gradient values. The measure used in this calculation, defined as the positive quantity g_r, is given by

$$g_r = \log_e \left\{ \sum_{i,j \subset R} thr(-G_r, 1) \right\} \tag{6}$$

where $thr(x,1)$ equals x, if $x > 1$ and equals 0 if $x < 1$. The quantity g_r is therefore the log of the sum of all those radial gradient values within region R that are more negative than (i.e., steeper than), a gradient of -1 mm/mm. (The log was taken because the resultant distribution of values more closely approximate a normal distribution.) We similarly calculated indices for gradients in the nasal and temporal halves of region R, denoting these by parameters g_r^N and g_r^T, respectively.

Third, we calculated an index of maximum cup depth, defined as the average of the 500 largest depth values, measured within region R. Denoting this average as I_{500}, we define the index as

$$z_{500} = \gamma I_{500} - z_0. \tag{7}$$

Fourth, the fit to the image of a curved surface without a cup (i.e., equation 1, but without the first term on the right-hand side) was calculated, by analogy with equation 2, and is denoted by f_p. This value is low in normal images, particularly those in which the cup is small or absent.

Classification

For a given set of D parameters measured from each image (the procedure for selecting these is described below), i.e., a data vector $\mathbf{x} = (x_1, x_2, \ldots x_D)$, we calculated the probability that the point came from the normal group—that is, $P(\mathbf{x}|\mathrm{N})$, and the probability that it came from the glaucoma group i.e., $P(\mathbf{x}|\mathrm{G})$. These probabilities were calculated using the multivariate normal probability density function[35]:

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\mathbf{C}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{u})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{u}) \right\} \tag{8}$$

where $\mathbf{u}$ is the mean of $\mathbf{x}$ taken over the group in question (normal or glaucomatous) and $\mathbf{C}$ is the within-group covariance matrix. When cross-validation was used, data from the case being classified were excluded from the data used to calculate the means and the covariance matrices. This gives a less biased estimate of the ability of the classification method to generalize to new data (data not used to derive the classification method in the first place).

We then calculated the probability that the measurements were from an eye with glaucoma, i.e., $P(\mathrm{G}|\mathbf{x})$. According to Bayes' theorem,[35] this is

$$P(G|\mathbf{x}) = \frac{P(\mathbf{x}|G)P(G)}{P(\mathbf{x}|N)P(N) + P(\mathbf{x}|G)P(G)} \tag{9}$$

where $P(\mathrm{N})$ and $P(\mathrm{G})$ are the prior probabilities that the subject in question is normal or has glaucoma. In the present case the prior probabilities were 0.5 because the sample sizes ($n = 100$) were equal, but this would not be the case with unequal sample sizes or if the normal population were screened.

Parameter Selection

Not all the described parameters were used for the purpose of classifying images. We excluded those that showed little difference between the groups (s, z_0 and b). Some sets of parameters were obviously closely related (e.g., z_m and z_{500}; f and f_R; and g_r, g_r^N, and g_r^T) and for these parameters we took the one showing the largest difference, measured by d', between the two groups. One parameter, horizontal image slope (a) was excluded because the difference between the two groups could have been artifactual (see the Discussion section). This resulted in a set of seven parameters, defined as $\mathbf{x} = c, d, z_{500}, g_r^T, f_p, f_R$, and r_0. These were used to calculate, for each image, the probability that it came from the glaucoma group—that is, $P(\mathrm{G}|\mathbf{x})$. Cases were classified as glaucoma if $P(\mathrm{G}|\mathbf{x}) > 0.5$. Because $P(\mathrm{G}|\mathbf{x}) = 1 - P(\mathrm{N}|\mathbf{x})$ this is equivalent to the condition that $P(\mathrm{G}|\mathbf{x}) > P(\mathrm{N}|\mathbf{x})$. With this method, specificity (the percentage of normal cases correctly classified) and sensitivity (the percentage of glaucoma cases correctly classified) tend to be similar and equal to the overall classification accuracy.