

# Automated Segmentation of the Optic Disc from Stereo Color Photographs Using Physiologically Plausible Features

Michael D. Abramoff,<sup>1,2,3</sup> Wallace L. M. Alward,<sup>1,2</sup> Emily C. Greenlee,<sup>1,2</sup> Lesya Shuba,<sup>1</sup> Chan Y. Kim,<sup>1</sup> John H. Fingert,<sup>1</sup> and Young H. Kwon<sup>1</sup>

**PURPOSE.** To evaluate a novel automated segmentation algorithm for cup-to-disc segmentation from stereo color photographs of patients with glaucoma for the measurement of glaucoma progression.

**METHODS.** Stereo color photographs of the optic disc were obtained by using a fixed stereo-base fundus camera in 58 eyes of 58 patients with suspected or open-angle glaucoma. Manual planimetry was performed by three glaucoma faculty members to delineate a reference standard rim and cup segmentation of all stereo pairs and by three glaucoma fellows as well. Pixel feature classification was evaluated on the stereo pairs and corresponding reference standard, by using feature computation based on simulation of photoreceptor color opponency and visual cortex simple and complex cells. An optimal subset of 12 features was used to segment all pixels in all stereo pairs, and the percentage of pixels assigned the correct class and linear cup-to-disc ratio (LCDR) estimates of the glaucoma fellows and the algorithm were compared to the reference standard.

**RESULTS.** The algorithm was able to assign cup, rim, and background correctly to 88% of all pixels. Correlations of the LCDR estimates of glaucoma fellows with those of the reference standard were 0.73 (95% CI, 0.58–0.83), 0.81 (95% CI, 0.70–0.89), and 0.86 (95% CI, 0.78–0.91), respectively, whereas the correlation of the algorithm with the reference standard was 0.93 (95% CI, 0.89–0.96;  $n = 58$ ).

**CONCLUSIONS.** The pixel feature classification algorithm allows objective segmentation of the optic disc from conventional color stereo photographs automatically without human input. The performance of the disc segmentation and LCDR calculation

of the algorithm was comparable to that of glaucoma fellows in training and is promising for objective evaluation of optic disc cupping. (*Invest Ophthalmol Vis Sci.* 2007;48:1665–1673) DOI:10.1167/iov.06-1081

Glaucoma is the second leading cause of blindness in the United States, characterized by gradual damage of the optic nerve and resultant visual field loss.<sup>1</sup> Early diagnosis and optimal treatment have been shown to minimize the risk of visual loss due to glaucoma.<sup>2</sup> The hallmark of glaucomatous progression is cupping of the optic disc. One way of determining the amount of cupping is planimetry performed by experienced glaucoma specialists examining stereo color photographs of the optic disc.<sup>3,4</sup>

Pixel feature classification is a machine learning technique that can be used to assign a class to the pixels in an image, in this case the cup, rim, or background.<sup>5–8</sup> Pixel feature classification uses multiple pixel features, which are numeric properties of a pixel and its surroundings. Originally, pixel intensity was used, and later, its contrast with the surrounding region, its proximity to an edge, and so forth, were added. As a supervised algorithm, it has two stages: a training stage, in which the algorithm “statistically learns” to classify pixels correctly from known classifications, and a testing or classification stage, in which the algorithm classifies a new, previously unseen image. The number of features for any pixel is essentially infinite, and it is easy to prove that one or more subsets of this infinite set are optimal for classifying the image according to some reference standard. Hundreds of features for a pixel can be calculated in the training stage to cast as wide a net as possible. Still, because all potential features cannot possibly be taken into account, such a selection is arbitrary to some extent and carries a bias. One of the advantages of the pixel feature classification method is that it avoids the potentially much larger bias of an imaging expert’s developing the algorithm by selecting one or two feature detectors. This is common in unsupervised image classification, such as in an edge-based, deformable model or in template-based algorithms that have been used in segmentation of retinal images and optic disc in nonstereo images.<sup>9</sup> Pixel feature classification has so far outperformed all other approaches for the segmentation of vessels, hemorrhages and microaneurysms, exudates, infarcts, and drusen in retinal images<sup>7,8</sup> (Niemeijer et al. *IOVS* 2005;45:ARVO E-Abstract 3468).

Another advantage of pixel feature classification is its potential for mimicking human experts’ visual performance by using parallel visual “channels.”<sup>10</sup> We have designed pixel features for color that mimic the output of the dark-bright, red-green, and blue-yellow center surround opponency cells, with different receptive field sizes and orientation, and features for stereo disparity that mimic the output of binocular cells in area V1 of the visual cortex.<sup>11,12</sup>

The purpose of this study was to compare the performance of pixel feature classification that segments the optic disc into cup, rim, and background from color stereo photographs and

From the <sup>1</sup>Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, Iowa; the <sup>2</sup>Department of Veterans Affairs, Iowa City VA Medical Center, Iowa City, Iowa; and the <sup>3</sup>Department of Electrical and Computer Engineering, University of Iowa, Iowa City, Iowa.

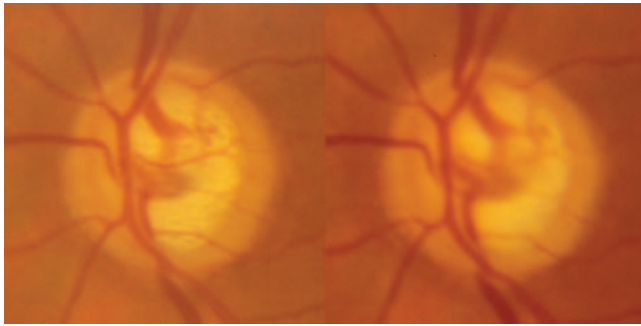
Supported by National Eye Institute Grant R01-EY017066, Research to Prevent Blindness, the Wellmark Foundation, Department of Defense Grant A064-032-0056, U.S. Department of Agriculture, University of Iowa, Netherlands Organization for Health Related Research (MDA), The National Eye Institute, a Lew R. Wasserman award from Research to Prevent Blindness (WLMA), and an unrestricted educational grant from Allergan Corp. (CYK).

Submitted for publication September 11, 2006; revised October 20, 2006; accepted January 23, 2007.

Disclosure: M.D. Abramoff (P); W.L.M. Alward, None; E.C. Greenlee, None; L. Shuba, None; C.Y. Kim, None; J.H. Fingert, None; Y.H. Kwon (P)

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Michael D. Abramoff, Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA 52242; michael-abramoff@uiowa.edu.



**FIGURE 1.** Stereo color image pair of the optic nerve head of the left eye after alignment and cropping to  $512 \times 512$  pixels. The left and right stereo images can be fused by holding the page at approximately 30 cm (12 in.) distance from the eye and gazing into infinity.

the performance of glaucoma fellows with the results of three glaucoma faculty experts, for measurement of glaucoma progression.

Specifically, we compared the performance of the pixel classification algorithm to segment the optic disc photographs into cup, rim, and background with the planimetric performance of the fellows against the reference standard planimetric segmentation by the glaucoma experts.

## METHODS

### Stereo Color Images of the Disc

The study was approved by the University of Iowa's Institutional Review Board and adhered to the tenets of the Declaration of Helsinki. For the test set, stereo photographs of the optic disc were obtained from 58 consecutive patients of the Glaucoma Service at the university. Patients were included if they met the definition of suspected glaucoma or open-angle glaucoma. Those with a history of angle closure or combined-mechanism glaucoma, or any nonglaucomatous optic neuropathy were excluded. The diagnosis of "suspected glaucoma" was based on a suspect optic nerve appearance (enlarged cupping on clinical examination) with normal visual field and intraocular pressure (IOP)  $\leq 21$  mm Hg. It also included those with normal optic disc appearance on biomicroscopy and normal visual field, but with elevated IOP ( $>21$  mm Hg). The "open-angle glaucoma" diagnosis included patients with both primary and secondary open-angle glaucoma and was based on the presence of an open iridocorneal angle, glaucomatous optic disc, and/or nerve fiber layer defects on biomicroscopy, and visual field changes (regardless of IOP level). Glaucomatous optic discs were identified as those with either diffuse or focal thinning of the neuroretinal rim. Visual field abnormalities were considered to be glaucomatous if they were consistent with the optic nerve examination and had either a (1) typical nerve fiber layer distribution or (2) glau-

coma hemifield test result outside normal limits. The diagnoses were made by the treating faculty glaucoma specialist according to these definitions.

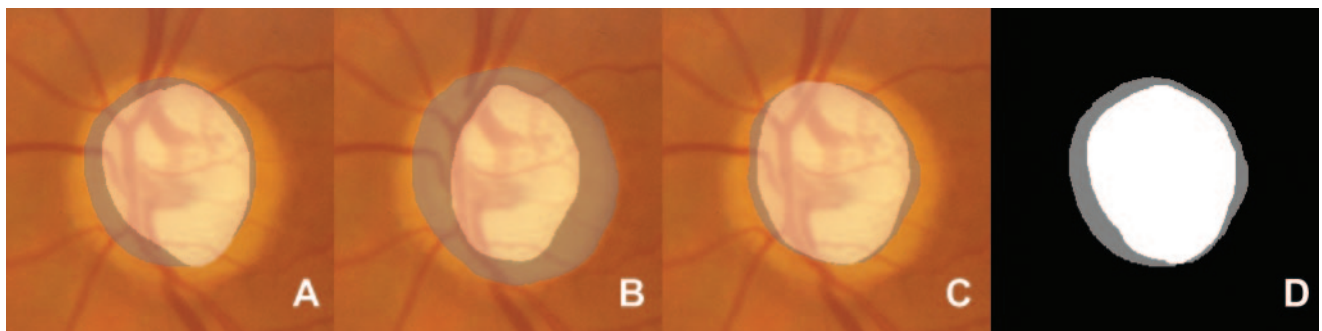
Color slide stereo photographs centered on the optic disc of both eyes were acquired with a fixed stereo-base retinal camera (3Dx; Nidek, Newark, NJ). After development, slides were scanned at a  $4096 \times 4096$ -pixel resolution, 24-bit depth with a slide scanner (Kodak Inc., Rochester, NY). Only one eye of each patient was selected for the study based on the quality of the stereo image. To limit the number of pixels ( $>16$  million for each left-right stereo pair) that would otherwise have to be processed by the algorithm, the left and right stereo images of the selected stereo pairs were cropped to  $512 \times 512$  pixels each with the optic disc in the center, by automatically locating it in the  $4096 \times 4096$  images and cropping to 256 pixels outward from the center of the disc based on its brightness.<sup>13</sup> The cropped images included the optic disc in its entirety in all images (see Fig. 1).

Computer-aided planimetry of the stereo pairs with a hand-held stereoscope was performed by three faculty glaucoma specialists at the University of Iowa on the 58 stereo pairs by using custom-made annotation software.<sup>14</sup> In addition, three glaucoma fellows in the university's Glaucoma Fellowship performed planimetry on all stereo pairs in the same manner. The human graders were masked from the patient identifier, and the clinical information was associated with the photographs and the results of all other graders at the time of grading. Graders were instructed to outline the disc margin while viewing the photographs with a stereo viewer. Then, they were instructed to draw the cup to the best of their ability. When the cup margin was difficult to identify (usually at the temporal slope of the disc), they were instructed to identify and draw the cup margin on the slope that is half way between the bottom of the cup and top of the rim.

A *reference standard* grading with a unique class (cup, rim, and background) assigned to each pixel was then created by combining the grading by each faculty grader in a "majority-win" manner: each pixel was assigned the class that received the majority of votes (e.g., if two graders voted a pixel to be rim and one voted it to be cup, the pixel assigned was rim). In the case of a draw, the pixel was assigned the background class (Figure 2).

### Pixel Feature Classification

**Rationale for Inclusion of Specific Features.** Originally, pixel intensity and simple edge operators were used as pixel features.<sup>5</sup> Once it was realized that a pixel's feature could also incorporate the adjacent pixels (i.e., its surround) and the Gaussian filter bank concept had been established mathematically,<sup>15</sup> it was possible to use filter bank outputs to create a larger set of features, from which an optimal combination of features could be selected—a method used initially in pulmonary image analysis.<sup>16</sup> Gaussian filter bank features are sensitive to edges and textures in the gray-scale intensity image obtained from a color image at different scales and orientations. For more optimal use of the color information in fundus color images, we added the outputs of a Gaussian steerable (i.e., having multiple orientations) filter bank in

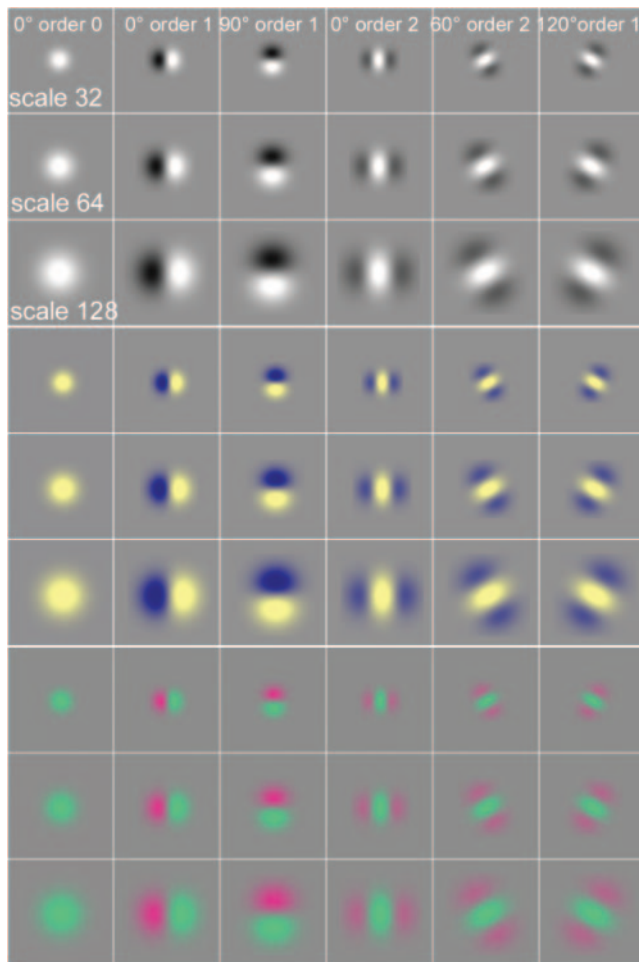


**FIGURE 2.** (A–C) Gradings of an optic disc stereo pair by three faculty glaucoma specialists. Rim is grayish, cup is whitish in the left image of the stereo pair from Figure 1. (D) Reference standard created from (A–C) with white, cup; gray, rim; and black, background

hue, saturation, brightness space as features and also the variance in the red, green, and blue channels in a small region around the pixel.<sup>8</sup>

Such features are obviously different from the color opponency model of primate color vision, and for this study we added the output of a Gaussian steerable filter bank in color opponency space for dark-bright, red-green, and blue-yellow center-surround color opponency. See Appendix A for the mathematical details. Preliminary experiments showed that the color opponency filter bank improved the performance (Figure 3).

Similarly, because expert human graders require stereo photographs for optimal segmentation of the optic nerve head, we decided to add horizontal stereo disparity maps as features. In the literature, many known algorithms use edge and texture information and some form of energy minimization to solve this depth from stereo problem in color or video images.<sup>17</sup> These algorithms are fundamentally different from what is known about stereo disparity computation in primates.<sup>18</sup> In preliminary studies, we found that the stereo disparity algorithms that have been found to be optimal for natural scene stereo disparity—namely, the graph search algorithms first described by Kolmogorov and Zabih<sup>17</sup> and Birchfield and Tomasi<sup>19</sup>—did not perform as



**FIGURE 3.** Example of the variety of color opponency, steerable, Gaussian filter bank kernels. Kernels are oriented as if the local gradient were horizontal. *First row, from left to right:* dark-bright opponency kernels for zeroth order, first order 0° to the local gradient, first-order 90° to the local gradient, second-order 0° to the local gradient, second order 60° to the local gradient, and second order 150° to the local gradient, at scale  $\sigma$  32 pixels. *Second row:* same, for scale  $\sigma$  64 pixels; *third row:* 128 pixels. Next *three rows:* same, for blue-yellow opponency kernels; *last three rows:* same, for red-green kernels. Smaller scales are not shown because they are difficult to see. These kernel images also represent the response of each of the features to an impulse function.

well as expected on the low-contrast/high-noise images in this study. Therefore, we decided to add features calculated by simulating the output of binocular complex cells in V1 visual cortex as first described by Qian and Zhu.<sup>12</sup> The output of binocular simple cells in the mammalian V1 has been shown to correspond closely with the convolution of its receptive field with Gabor-basis kernels, with the same spatial frequency and scale but different phases for the left and right fields. Stereo disparity-sensitive complex cells can then be modeled by summing the squared outputs of two simple cells with different phases, but the same size receptive field. See Appendix B for the mathematical background.

An informal taxonomy of pixel features obtainable from stereo color images may be:

- “Less physiologically plausible”: The outputs of edge detectors such as those described by Canny and Sobel, or graph search stereo disparity algorithms, as there is currently no evidence that these are expressed by any neural circuit.<sup>20</sup>
- “Somewhat physiologically plausible”: The outputs of Gaussian filter banks in intensity, or RGB, space, because there is evidence that a Gaussian profile is expressed by certain neural circuits.<sup>21</sup> No red/green/blue- or hue/saturation/intensity-sensitive receptive fields have been described to our knowledge.
- “More physiologically plausible”: Gabor wavelets in color opponency space, and the output of simulations of binocular complex cells, as there is evidence that these are expressed in certain neural circuits.<sup>22,23</sup>

We have assumed that a combination of retinal and cortical image processing as found in biological systems would be effective in classifying the stereo color images of the optic disc and thus have implemented these physiologically plausible features and added them to the set of potential features from which an optimal set of features is then selected. Whether these features indeed increase the performance of the system can be studied (implicitly) by running feature selection on sets with and without the “more physiologically plausible” features and comparing the accuracy.

### A Priori Expectation Feature

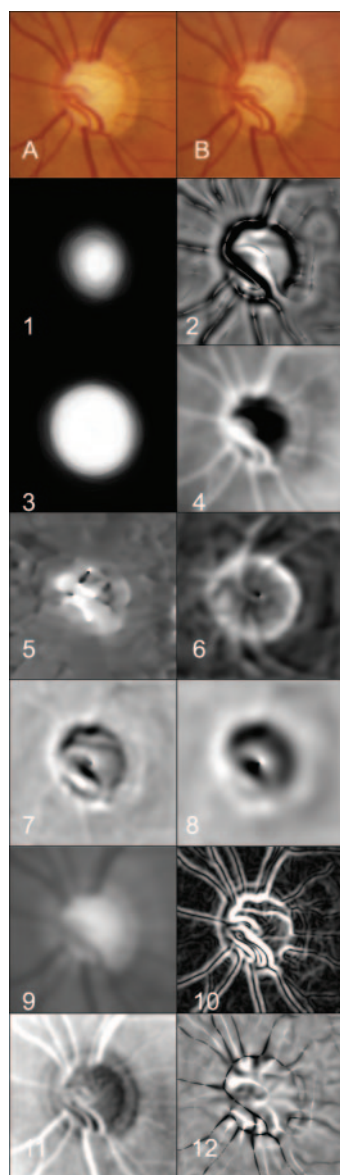
In pixel feature classification, non pixel-based information is not naturally included. However, it is easy to see that in stereo disc color imaging of this type, the likelihood that a pixel in the center of the image is cup is higher than the likelihood that it is rim or background. To capture this a priori expectation of the class of a pixel based on its location in the image, an a priori rim-nonrim feature was calculated by performing principal component analysis on all reference standard images and taking the first principal component.<sup>24</sup> This feature expresses the expectation of any pixel’s being rim or either background or cup based on its position. Preliminary studies without the a priori feature did not attain an accuracy of more than 73% (see the Results section).

A total of  $108 + 3 \times 48 + 1$  features were thus calculated, for all 15,204,352 pixels in all 58 color stereo pairs. Because these features can be calculated independently in parallel, as is the case in the primate visual system, this is a naturally parallel computation. See Figure 4 for an example of all values of the 12 optimal pixel features that were selected for a specific stereo pair.

### Classifier Implementation

In pixel feature classification, a classifier is trained on a subset of all data, the training set, and then repeatedly tested on another subset of the data, the test set. By repeatedly testing with different subsets of all the features and comparing the performance of the classifier with these different feature subsets, the optimal subset of features can be selected. In preliminary studies, the k-nearest neighbor (kNN) classifier outperformed support vector machines and linear discriminant analyzer classifiers, as has been found in other studies.<sup>8,16,25,26</sup> A kNN classifier has the advantage that no elaborate training is required and





**FIGURE 4.** Stereo pair and its most discriminant pixel features. (A, B) The left and right stereo image of the stereo pair. Rows below them show images of the most dominant 12 features in rank order (see Table 1 for a description of these features).

that only one parameter needs to be adjusted, the  $k$ . We have interfaced the publicly available kNN library ([www.cs.umd.edu/~mount/ANN/](http://www.cs.umd.edu/~mount/ANN/)) developed by David Mount and Sunil Arya and provided in the public domain by the Department of Computer Science, University of Maryland, College, Park, MD) to Java.<sup>27</sup>

## Data Analysis

The classifier was optimized on the accuracy as explained in Appendix C to find the set of optimally discriminant features. A clinically more relevant measure was also used to compare the performance of the pixel feature classification and the three glaucoma fellows to the reference standard. For this purpose, all pixels in all stereo pairs were classified in a leave-one-out fashion.<sup>25</sup> The resultant gradings contained the algorithm's estimate for each pixel as to whether that pixel was cup, rim, or background, similar to the gradings by the glaucoma faculty and the glaucoma fellows. A linear cup-to-disc ratio was then computed for each graded image as a square root of area cup-to-disc ratio:

$$cd_{r,linear} = \sqrt{\frac{n_c}{n_r + n_c}},$$

where  $n_c$  is the number of pixels in cup, and  $n_r$  is the number of pixels in rim.<sup>28</sup>

## Optimal Feature Selection

The algorithm had been trained on a larger set of 101 stereo disc pairs of patients with glaucoma (Abràmoff et al., *IOVS* 2006;46:ARVO E-Abstract 3635). The training set was used to determine which combination of features gives the best accuracy ("feature selection," see Appendix C). A hard classification using "majority-win" of the kNNs was used as the classification for each pixel. Each feature column in the training set was scaled to have zero mean and unit variance over all pixels. The most discriminant features were then selected in a sequential forward-floating search,<sup>25</sup> by adding all features to an initially empty feature set one by one and retaining only the new feature that showed the highest gain in classification performance. Preliminary studies showed that  $k = 501$  gave slightly better performance over a range of feature vector lengths, but sensitivity to changing from 1 to 1001 was low. The number of samples was 1200 per image, per class. The pixel feature classification was implemented in Java, and it took 15 minutes to calculate 12 features and completely grade a stereo pair on a standard 2-GHz 2-Gbyte computer (running Windows XP; Microsoft, Redmond, WA).

Table 1 shows the 12 features that were found to be most dominant. Figure 5 shows an example stereo pair and those pixel features that were most discriminant features. After feature selection, the algorithm was ready for testing using this optimal set of features.

## RESULTS

The stereo pairs of 58 eyes of 58 patients were included to test the algorithm. The age of these patients was  $61.2 \pm 17.8$  years (mean  $\pm$  SD), and there were 33 women. There were 22 patients with suspected glaucoma and 36 with open-angle glaucoma. Of the total of 58 eyes, 27 underwent Humphrey visual field testing (24-2 SITA Standard, HFA II; Carl Zeiss Meditec, Dublin, CA) within 1 year of the disc photographs. The remainder ( $n = 31$ ) had either Goldmann visual field testing ( $n = 28$ , Goldmann Perimeter 940; Haag Streit, Liebefeld, Switzerland) or no field tests ( $n = 3$ ). We performed Goldmann visual field tests in patients whose Humphrey visual field result was not reliable and/or who had moderate to severe visual field defect (typically  $-15$  dB mean deviation [MD] or worse). In these patients, we found the Goldmann visual field results to be more reliable. A few patients ( $n = 3$ ) were unable to perform any visual field tests at all (e.g., those with Alzheimer's disease). The Humphrey visual field mean defect (in decibels  $\pm$  SD) in eyes with suspected glaucoma ( $n = 11$ ) or open-angle glaucoma ( $n = 16$ ) were  $-1.1 \pm 2.3$  and  $-4.1 \pm 5.2$ , respectively. The Humphrey visual field pattern SD (dB  $\pm$  SD) in eyes with suspected glaucoma and open-angle glaucoma was  $2.8 \pm 3.6$  and  $4.5 \pm 4.5$ , respectively. The Goldmann visual field results of the suspected-glaucoma group ( $n = 8$ ) were within normal limits, but the results in the open-angle group ( $n = 20$ ) showed a variety of defects consistent with glaucomatous nerve fiber layer loss including paracentral scotoma, nasal step, and arcuate defect(s).

Average disc size (cup and rim area, as determined from the reference standard) was  $14,085 \pm 3,467$  (SD) pixels. All pixels in all stereo pairs were graded by the algorithm. Accuracy was 0.876; the algorithm was able to assign 88% of all pixels in the 58 stereo pairs to the correct class. Figure 5 shows how the gradings by the glaucoma faculty, the glaucoma fellows, and the algorithm compared with each other visually. Correlations of the linear cup to disc ratio for each of the glaucoma fellows with the reference standard were 0.73 (95% confidence inter-

TABLE 1. Twelve Features Selected as the Most Dominant

Rank	Description	Scale (Pixels)	Cumulative Accuracy on Training Set
1	A priori probability of pixel being cup	1	0.810
2	Red-green opponency (opp.) second-order Gaussian derivative transversal to local gradient	32	0.825
3	A priori probability of a pixel being rim	1	0.829
4	Red-green opp. Gaussian blurred (zeroth order)	8	0.832
5	Complex cell output tuned to a disparity of −12 pixels	16	0.845
6	Blue-yellow opp. second-order Gaussian derivative	32	0.854
7	Red-green opp. first-order Gaussian derivative perpendicular to local gradient	16	0.869
8	Bright dark opp. Gaussian blurred (zeroth order)	16	0.872
9	Red-green opp. first-order Gaussian derivative radial to local gradient	4	0.874
10	Blue-yellow Gaussian blurred (first-order)	4	0.874
11	Red-green opp. first-order Gaussian derivative parallel to local gradient	4	0.875
12	Red-green opp. Gaussian blurred (zeroth order)	32	0.876

The features are ranked by their contribution to cumulative accuracy on the training set.

val [CI], 0.58–0.83), 0.81 (95% CI, 0.70–0.89), and 0.86 (95% CI, 0.78–0.91), respectively, whereas correlation of the algorithm with the reference standard was 0.93 (95% CI, 0.89–0.96;  $n = 58$ ).

Subgroup analysis showed no significant differences in the correlation between the suspected and open-angle glaucoma groups for each of the fellows or the algorithm, and there was no dependency on disc size ( $P > 0.05$ ). Figure 6 shows scatterplots of the estimates for the linear cup-to-disc ratio (CDR) for each stereo pair for the glaucoma fellows and the algorithm versus the reference standard.

DISCUSSION

Our results show that pixel feature classification with biologically inspired features can classify a set of optic disc stereo

images with an accuracy that is comparable to that of glaucoma fellows. Other studies have also shown agreement of pixel classification algorithms for the segmentation of retinal vessels and the detection of diabetic retinopathy lesions with classifications made by human experts.<sup>8,26</sup> In fact, Niemeijer et al.,<sup>7</sup> found that a pixel classification-based approach for retinal vessel segmentation outperformed other supervised and unsupervised methods. The same group of investigators used a pixel classification approach to detect red diabetic retinopathy lesions and reached a sensitivity of 100%, a specificity of 87%, and an area under the ROC curve of 98.5% in a study of 100 patients.<sup>8</sup>

We are aware of only one previous study of classification of the optic disc in which stereo photographs were used.<sup>29</sup> In that study, a matching cost approach was used, based on pyramidal cross correlation. Recent depth from stereo algo-

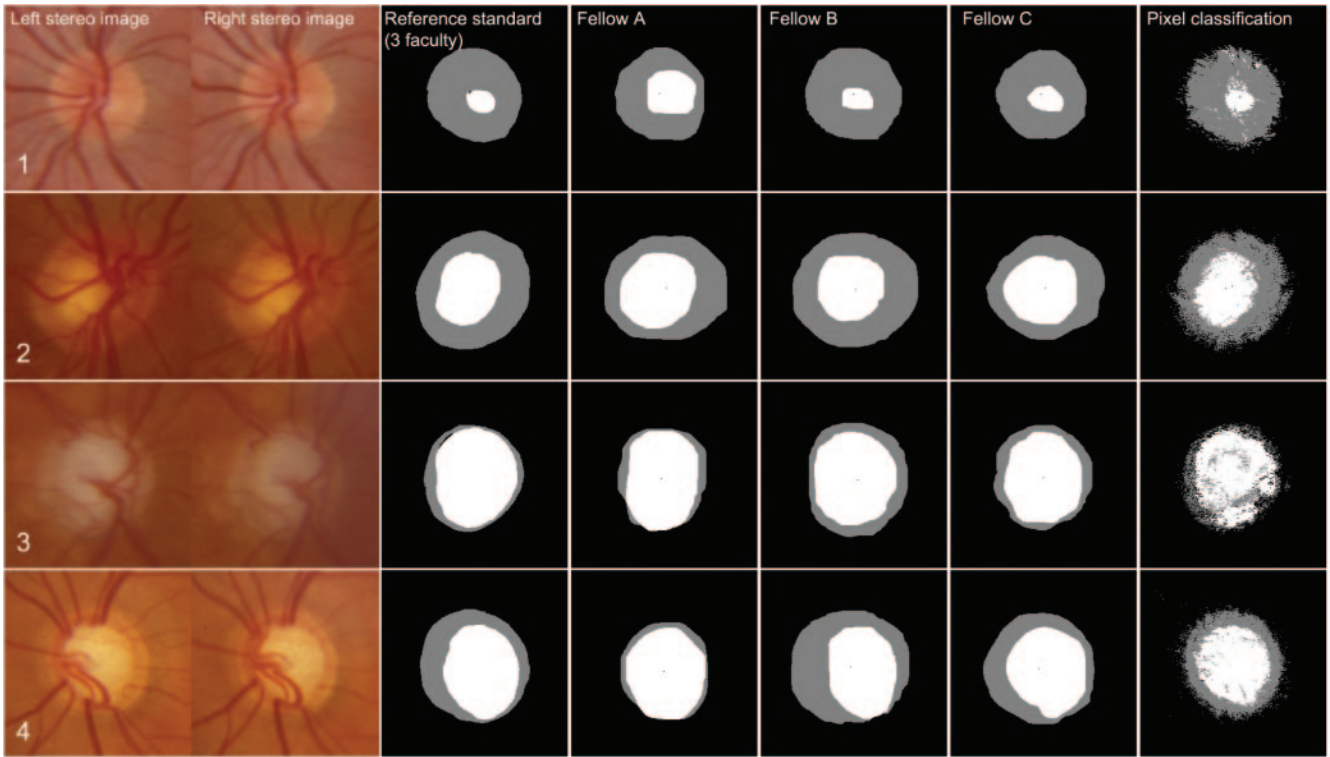
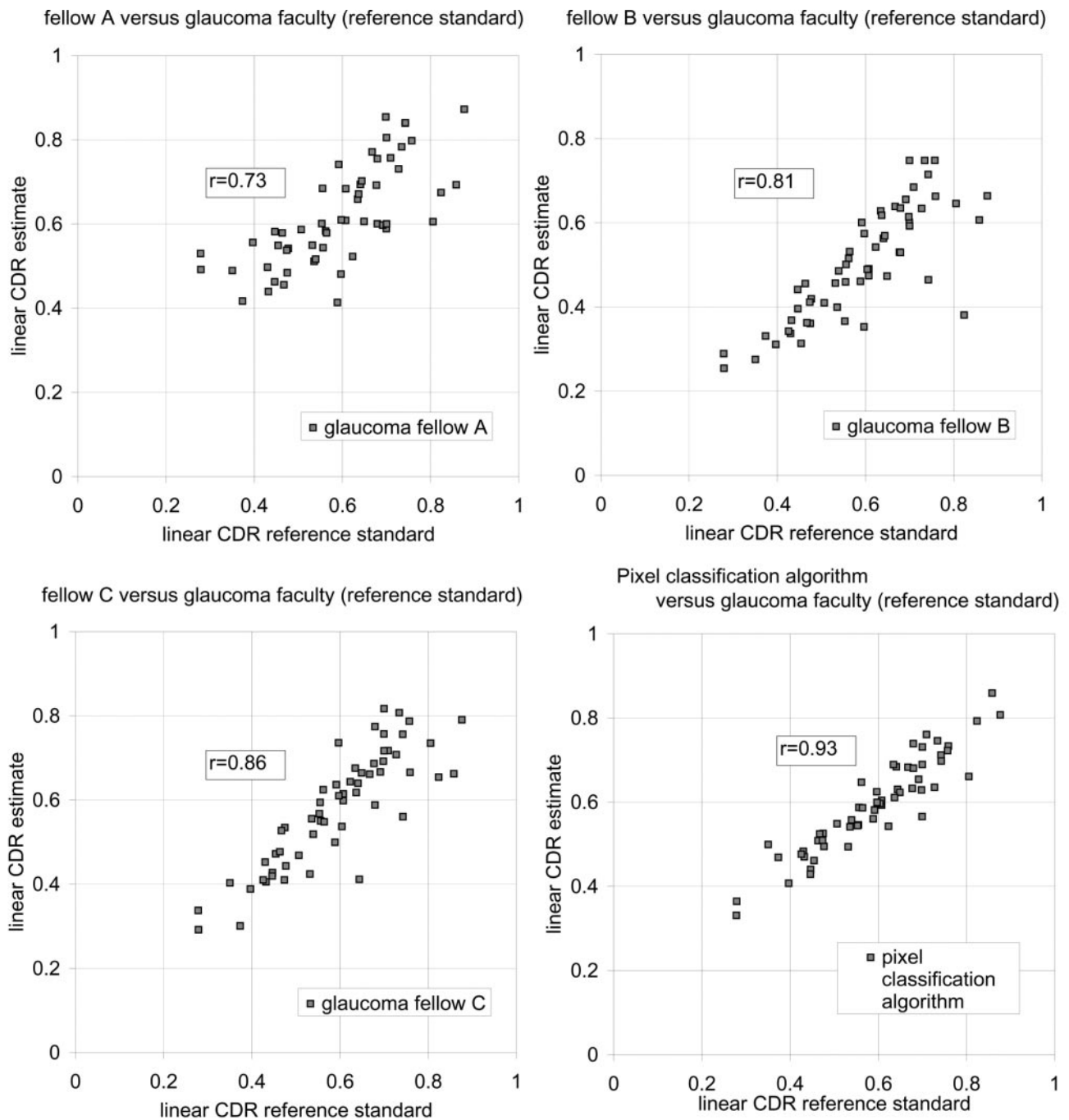


FIGURE 5. Classification of stereo pairs by glaucoma faculty, three glaucoma fellows, and pixel feature classification. Rows 1–4: small, medium, and large disc excavation and excavation with inferior notching, respectively. From left to right: left and right images of the stereo pair; reference standard by three glaucoma expert faculty; grading by fellow A, fellow B, and fellow C; and grading by pixel feature classification.



**FIGURE 6.** Scatterplots of linear cup-to-disc ratio estimates from all 58 stereo pairs by three glaucoma fellows and the pixel feature classification algorithm (vertical axes), against the reference standard grading by three glaucoma faculty (horizontal axes). Clockwise, fellows A and B, the algorithm, and fellow C respectively.  $r$ , is the correlation coefficient for each of these estimates with the reference standard.

gorithms, mostly based on graph searches, have been shown to outperform matching cost computation algorithms on real-world depth scenes.<sup>17,19,30</sup> As shown by our results, these graph search algorithms in their turn can be outperformed by an algorithm that mimics binocular cells in the visual cortex in optic disc stereo images.<sup>12</sup>

On a computational level, the role of most features is to determine how much a patch around the pixel resembles the filter kernel of that feature. The results indicate that the features selected for optimal accuracy of the system were predominantly of the "somewhat" and "more" physiologically plausible categories, and not for example, the familiar Sobel or

Canny intensity edge detectors. On an image-understanding level, it is difficult to explain the role of the selected features. A second-order Gaussian derivative of a specific orientation and blue-yellow color opponency can be said to correspond to "striped oriented patches" of the pixel surround, but is more than a simple edge detector, and also different from a texture detector. This situation is analogous to the difficulty in describing the role of different receptive field properties for scene analysis by the primate visual cortex.

Objective classification of optic disc cup and rim from color stereo photographs is promising, because it allows optic disc stereo photographs to be evaluated objectively and rapidly.



However, the ultimate goal of an algorithm such as this is the objective measurement of glaucoma progression, or enlargement of the optic cup, over time. The results of the present study are intended foremost to serve as a proof of concept, and many aspects can be improved, including adding features that mimic physiology even more and adding vessel segmentation.

Nevertheless, several serious issues remain to be addressed before the present algorithm can be thus used. First, the algorithm is deterministic and therefore returns the same classification if offered the same stereo pair repeatedly. However, whether it also gives very similar classifications on two stereo pairs of the same patient of the same eye taken minutes apart, with subsequent patient movement, vignetting, and refocusing, has not been determined.

Second, there is room for substantial improvement, as is evident from the noisy visual appearance of the classification, compared with the smooth boundaries of the reference standard, and the percentage of correctly classified pixels, which is only 88%. For example, compared with the optic disc images obtained with a fundus camera (Carl Zeiss Meditec), the 3Dx (Nidek) gives decidedly darker and less crisp images, which may have affected the quality of the performance negatively. We have adopted several different postprocessing techniques that are showing promising results.<sup>31</sup>

Third, the algorithm performs well on stereo pairs obtained with the fixed-base 3Dx, which may not be widely available, and it has not been established whether the algorithm will perform similarly on non-fixed-base stereo pair cameras.

Fourth, the algorithm's performance is highly dependent on a single feature, the prior PCA. One of its roles is likely to be to indicate that the disc is in the center of the image, and the cup is in the center of the disc and surrounded by the rim. Because this is an assumption about the expected size and location of the disc, the algorithm's accuracy may be lower on discs that are smaller or larger than in our sample and also if the disc is not properly centered in the cropped image. Therefore, we are currently engaged in removing this feature from the selected features while maintaining performance.

Finally, the algorithm is slow. It is important to understand that the current implementation of the algorithm was not written for speed, but for ease of use and modifiability, and that it can easily be optimized for speed. In our previous optimized algorithms, classification times of 15 seconds for a whole image were reached.<sup>8</sup> In addition, calculation of the features and feature classification is naturally parallel, so that with appropriate hardware, the features for all pixels can be calculated in parallel.

Contrary to expectations, stereo disparity was not the most discriminant feature for the performance of the algorithm. In Figure 1, it can be appreciated that the stereo photographs are not very crisp. Possibly, using a pair of photographs improves the quality of the fused image by suppressing these irregularities to some extent. In addition, human graders may be unaware of the relative importance of color and texture of the tissues compared with depth-from-stereo as the feature they use to segment the optic disc. The algorithm may therefore also perform reasonably on non-stereo color photographs.

In summary, the results of this study have shown that pixel feature classification with biologically inspired features of stereo color images is a good starting point for the classification of the optic disc and other three-dimensional structures on the retina, and also, that features based on the physiological process outperform standard pixel features on segmentation of the optic disc.

## APPENDIX A

### Color Opponency Gaussian Filter Bank

By careful selection of steerable Gaussian kernels in a Gaussian filter bank, second-order invariant features can be calculated, so that all other features up to the second order are combinations of the features in this invariant set.<sup>15</sup> Color opponency Gaussian filter banks<sup>32</sup>  $L_{db}, L_{rg}, L_{by}$ , for dark-bright (db), red-green (rg), and blue-yellow (by) opponency images, are computed from each color image as follows:

$$L_{db} = \frac{L_r + L_g + L_b}{3},$$

$$L_{rg} = L_r - L_g,$$

and

$$L_{by} = (L_r + L_g) - 2L_b,$$

with  $L_r, L_g, L_b$  the red (r), green (g), and blue (b) channels of the left ( $L$ ) image of the stereo pair, respectively.

$L_{db}, L_{rg}, L_{by}$  are then each convolved with Gaussian filter bank kernels to obtain the following color-opponency Gaussian derivative features:

$$L_0(\sigma), L_1^{0^\circ}(\sigma), L_1^{90^\circ}(\sigma), L_2^{0^\circ}(\sigma), L_2^{60^\circ}(\sigma), L_2^{120^\circ}(\sigma),$$

with  $L_{n,op}^\alpha(\sigma)$  the feature obtained by computing  $L_n^\alpha(\sigma) = L_1 \otimes G_n^\alpha(\sigma)$  for order  $n$ , orientation  $\alpha$  ( $0^\circ$  and  $90^\circ$  orientation to the local gradient for first-order derivatives and  $0^\circ$ ,  $60^\circ$ , and  $120^\circ$  to the local gradient for second-order derivatives), scale  $\sigma \in (2, 4, 8, 16, 32, 64)$ , and opponency (op) image  $L_{op} \in (L_{db}, L_{rg}, L_{by})$ .  $\otimes$  denotes convolution.  $L_n^\alpha$  is computed as follows:

$$L_0 = L_1$$

$$L_1^\alpha = \cos\alpha L_{xx} + \sin\alpha L_x$$

$$L_2^\alpha = \cos^2\alpha L_{xx} + \cos\alpha \sin\alpha L_{xy} + \sin^2\alpha L_{yy},$$

where the  $xx, yy$ , and  $xy$  subscripts to  $L$  denote differentiation of  $L_{op}$  to  $x$  and/or  $y$ <sup>15</sup> and results in 108 color opponency Gaussian filter bank features.

## APPENDIX B

### Stereo Features

Gabor features  $G_{\sigma,op}^\alpha$  were calculated by convolution of the three opponency images  $op \in (L_{db}, L_{rg}, L_{by})$  with Gabor wavelet kernels at scales  $\sigma = 1, 2, 4, 8, 16$ , and 32 pixels, at orientation  $\alpha = 0^\circ, 45^\circ, 90^\circ$ , and  $135^\circ$ :

$$G_{\sigma,op}^\alpha(x,y) = e^{-\frac{x^2}{2\sigma^2}} \cos(\omega x + \varphi),$$

with phase  $\varphi$  0 and with spatial frequency  $\omega = \frac{\pi}{\sigma}$  constant, for a total of 82 features. See Figure 3 for an example of these kernels.

Stereo disparity features were computed by first convolving the left and right color opponency images with left and right Gabor kernels  $k_l$  and  $k_r$ :

$$k_l(x) = e^{-\frac{x^2}{2\sigma^2}} \cos(\omega x + \varphi_l)$$

$$k_r(x) = e^{-\frac{x^2}{2\sigma^2}} \cos(\omega x + \varphi_r),$$

where  $\varphi_l, \varphi_r$  are the phases of the left and right Gabor wavelets, and  $\omega$  its preferred spatial frequency, which was kept constant at  $\omega = \frac{\pi}{\sigma}$ . The linear combination of convolutions of the left and right images with these respective kernels mimics the output of a simple cell:

$$S_\omega(\sigma, \varphi_l, \varphi_r) = I_{l,1} \otimes k_l + I_{l,r} \otimes k_r,$$

with  $I_{l,1}, I_{l,r}$  the left and right intensity images, respectively. The sum of the squares of two

$$S_{1,\omega}(\sigma, \varphi_l, \varphi_r), S_{2,\omega}\left(\sigma, \varphi_l + \frac{\pi}{2}, \varphi_r + \frac{\pi}{2}\right)$$

formed quadrature pairs:

$$C_\omega(\sigma, \Delta\varphi) = (S_{1,\omega}(\sigma, \varphi_l, \varphi_r))^2 + \left[S_{2,\omega}\left(\sigma, \varphi_l + \frac{\pi}{2}, \varphi_r + \frac{\pi}{2}\right)\right]^2,$$

with  $\Delta\varphi$  the phase difference

$$\Delta\varphi = \varphi_r - \varphi_l$$

between the right and left eye receptive fields. With  $\Delta\varphi \in \left\{ \pi \frac{6}{8}, \pi \frac{4}{8}, \pi \frac{2}{8}, 0, -\frac{2}{8}\pi, -\frac{4}{8}\pi, -\frac{6}{8}\pi \right\}$ ,  $C_\omega(\sigma, \Delta\varphi)$  was then tuned to stereo disparities of  $\{-4, -3, -2, -1, 0, 1, 2, 3\}$  pixels, respectively, calculated at scales  $\sigma = 1, 2, 4, 8, 16$ , and 32 pixels, and for each of the  $(L_{op}, R_{op}) \in \{(L_{db}, R_{db}), (L_{rg}, R_{rg}), (L_{by}, R_{by})\}$  color opponency pairs, resulting in 48 features for each of the color opponency spaces.<sup>12</sup>

## APPENDIX C

### Balanced Accuracy

Performance was calculated in the form of *balanced* accuracy (a combination of sensitivity and specificity in a single number) by calculating the number of correctly classified pixels in equally sized samples from the rim, the cup, and the background

$$\pi = \frac{N_{r,T} + N_{c,T} + N_{b,T}}{N_r + N_c + N_b},$$

with  $N_{r,T}$  the number of pixels correctly classified in a sample of rim pixels,  $N_{c,T}$  the number of pixels classified correctly in a sample of cup pixels,  $N_{b,T}$  the number pixels classified correctly as background in a sample of cup pixels, and  $N_r = N_c = N_b$  the total number of pixels in each sample. The reason to make this calculation is that otherwise, because of the predominance of background pixels in the image, an algorithm that is “good” in classifying background but “bad” in classifying rim would have an unfair advantage.

## References

1. Tielsch JM, Sommer A, Katz J, Royall RM, Quigley HA, Javitt J. Racial variations in the prevalence of primary open-angle glaucoma. The Baltimore Eye Survey. *JAMA*. 1991;266:369–374.
2. Heijl A, Leske MC, Bengtsson B, Hyman L, Bengtsson B, Hussein M. Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch Ophthalmol*. 2002;120:1268–1279.
3. Greaney MJ, Hoffman DC, Garway-Heath DF, Nakla M, Coleman AL, Caprioli J. Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma. *Invest Ophthalmol Vis Sci*. 2002;43:140–145.
4. Tielsch JM, Katz J, Quigley HA, Miller NR, Sommer A. Intraobserver and interobserver agreement in measurement of optic disc characteristics. *Ophthalmology*. 1988;95:350–356.
5. Panda DP, Rosenfeld A. Image segmentation by pixel classification in (grey level, edge value) space. *IEEE Trans Comp*. 1978;27:875–879.
6. Spence CD, Sajda P. Role of feature selection in building pattern recognizers for computer-aided diagnosis. *Proc SPIE*. 2006;3338:1434–1441.
7. Niemeijer M, Staal JS, van Ginneken B, Loog M, Abràmoff MD. Comparative study of retinal vessel segmentation on a new publicly available database. *Proc SPIE*. 2004;5370: 648–656.
8. Niemeijer M, van Ginneken B, Staal J, Suttorp-Schulten MS, Abràmoff MD. Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans Med Imag*. 2005;24:584–592.
9. Lowell J, Hunter A, Steel D, et al. Optic nerve head segmentation. *IEEE Trans Med Imag*. 2004;23:256–264.
10. Tailor DR, Finkel LH, Buchsbaum G. Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Res*. 2000;40:2671–2676.
11. Serre T, Riesenhuber M. Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. Cambridge, MA: Massachusetts Institute of Technology; CBCL Paper 239/AI Memo 2004-004. July 2, 2004.
12. Qian N, Zhu Y. Physiological computation of binocular disparity. *Vision Res*. 1997;37:1811–1827.
13. Abràmoff MD, Niemeijer M. Automatic detection of the optic disc location in retinal images using optic disc location regression. *Proceedings of the IEEE EMB Conference on Medical Imaging, New York, NY, August 30–September 3, 2006*. New York: Springer-Verlag; 2006:4432–4435.
14. Kwon YH, Kim YI, Pereira ML, Montague PR, Zimmerman MB, Alward WL. Rate of optic disc cup progression in treated primary open-angle glaucoma. *J Glaucoma*. 2003;12:409–416.
15. ter Haar Romeny BM. *Front End Vision and Multi-scale Image Analysis*. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2003.
16. van Ginneken B, Haar Romeny BM. Automatic segmentation of lung fields in chest radiographs. *Med Phys*. 2000;27:2445–2455.
17. Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts? *IEEE Trans Pattern Anal Mach Intell*. 2004;26:147–159.
18. Marr D. *Vision*. Freeman and Company, New York; 2005.
19. Birchfield S, Tomasi C. Depth discontinuities by pixel-to-pixel stereo. *Proceedings of the Sixth International Conference on Computer Vision, Bombay, India*. Washington, DC: IEEE Computer Society; 1998:1073–1080.
20. Sonka M, Fitzpatrick JM. *Medical Image Processing and Analysis*. Wellingham, WA: The International Society for Optical Engineering Press; 2000. *Handbook of Medical Imaging*, vol 2.
21. Segev R, Goodhouse J, Puchalla J, Berry MJ. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nat Neurosci*. 2004;7:1154–1161.
22. Lee TW, Wachtler T, Sejnowski TJ. Color opponency is an efficient representation of spectral properties in natural scenes. *Vision Res*. 2002;42:2095–2103.
23. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996;381:607–609.



24. Abramoff MD, Kwon YH, Ts'o DY, Li H, Barriga ES, Kardon R. A spatial truncation approach to the analysis of optical imaging of the retina in humans and cats. *Proc IEEE Int Symp Biomed Imag.* 2004;2:1115-1118.
25. Duda RA, Hart PE, Stork DG. *Pattern Classification*. New York: Wiley-Interscience, New York; 2001.
26. Staal JS, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B. Ridge based vessel segmentation in color images of the retina. *IEEE Trans Med Imag.* 2004;23:501-509.
27. Arya S, Mount D, Netanyahu N, Silverman R, Wu A. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *JACM.* 1998;45:891-923.
28. Parkin B, Shuttleworth G, Costen M, Davison C. A comparison of stereoscopic and monoscopic evaluation of optic disc topography using a digital optic disc stereo camera. *Br J Ophthalmol.* 2001; 85:1347-1351.
29. Corona E, Mitra S, Wilson M, Krile T, Kwon YH, Soliz P. Digital stereo image analyzer for generating automated 3-D measures of optic disc deformation in glaucoma. *IEEE Trans Med Imaging.* 2002;21:1244-1253.
30. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *ICJV.* 2002;47:7-42.
31. Merickel MB Jr, Wu X, Sonka M, Abramoff M. Optimal segmentation of the optic nerve head from stereo retinal images. *Proc SPIE.* 2006;6143; 61433B.
32. Calkins DJ, Tsukamoto Y, Sterling P. Microcircuitry and mosaic of a blue-yellow ganglion cell in the primate retina. *J Neurosci.* 1998;18:3373-3385.