Utility of Digital Stereo Images for Optic Disc Evaluation

Richard A. Stone,[1] Gui-shuang Ying,[1] Denise J. Pearson,[1] Mayank Bansal,[2] Manika Puri,[2] Eydie Miller,[1] Judith Alexander,[1,3] Jody Piltz-Seymour,[4] William Nyberg,[1] Maureen G. Maguire,[1] Jayan Eledath,[2] and Harpreet Sawhney[2]

Purpose. To assess the suitability of digital stereo images for optic disc evaluations in glaucoma.

Methods. Stereo color optic disc images in both digital and 35-mm slide film formats were acquired contemporaneously from 29 subjects with various cup-to-disc ratios (range, 0.26–0.76; median, 0.475). Using a grading scale designed to assess image quality, the ease of visualizing optic disc features important for glaucoma diagnosis, and the comparative diameters of the optic disc cup, experienced observers separately compared the primary digital stereo images to each subject's 35-mm slides, to scanned images of the same 35-mm slides, and to grayscale conversions of the digital images. Statistical analysis accounted for multiple gradings and comparisons and also assessed image formats under monoscopic viewing.

Results. Overall, the quality of primary digital color images was judged superior to that of 35-mm slides ($P < 0.001$), including improved stereo ($P < 0.001$), but the primary digital color images were mostly equivalent to the scanned digitized images of the same slides. Color seemingly added little to grayscale optic disc images, except that peripapillary atrophy was best seen in color ($P < 0.0001$); both the nerve fiber layer ($P < 0.0001$) and the paths of blood vessels on the optic disc ($P < 0.0001$) were best seen in grayscale. The preference for digital over film images was maintained under monoscopic viewing conditions.

Conclusions. Digital stereo optic disc images are useful for evaluating the optic disc in glaucoma and allow the application of advanced image processing applications. Grayscale images, by providing luminance distinct from color, may be informative for assessing certain features. (*Invest Ophthalmol Vis Sci.* 2010;51:5667–5674) DOI:10.1167/iovs.09-4999

From the [1]Department of Ophthalmology, University of Pennsylvania, Scheie Eye Institute, Philadelphia, Pennsylvania; [2]Vision Technologies, Sarnoff Corporation, Princeton, NJ; and the [4]Glaucoma Care Center, PC, Narberth, Pennsylvania.

[3]Present affiliation: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

Supported by National Institutes of Health (NIH) Bioengineering Research Partnership Grant R01-EY-017299-01, the Paul and Evanina Bell Mackall Foundation Trust (RAS), and Research to Prevent Blindness (RAS).

Submitted for publication November 30, 2009; revised April 5 and May 11, 2010; accepted May 13, 2010.

Disclosure: **R.A. Stone**, None; **G. Ying**, None; **D.J. Pearson**, None; **M. Bansal**, Sarnoff Corporation (E); **M. Puri**, Sarnoff Corporation (E); **E. Miller**, None; **J. Alexander**, None; **J. Piltz-Seymour**. None; **W. Nyberg**, None; **M.G. Maguire**, None; **J. Eledath**, Sarnoff Corporation (E); **H. Sawhney**, Sarnoff Corporation (E)

Corresponding author: Richard A. Stone, Department of Ophthalmology, University of Pennsylvania School of Medicine, 3535 Market Street, Suite 700, Philadelphia, PA 19104-3309; stone@mail.med.upenn.edu.

Glaucoma diagnosis, management, and research require complex assessments of the optic disc. Over many decades, the gold standard of optic disc evaluation in glaucoma has been the qualitative evaluation of stereo photographic images of the optic disc, most typically as 35-mm slides or other film-based media. More recently, automated optic disc analyzers have been introduced that use either laser technologies or polarized light to image the optic disc and/or the retinal nerve fiber layer. Studies with these instruments infrequently compare them directly to stereo photographs, and available reports generally find their diagnostic precision comparable to stereo photographs.[1–5] So far, the qualitative evaluation of stereo optic disc images by experienced observers remains the reference standard for optic disc evaluation in glaucoma[6–8] despite the subjectivity and variability inherent in this conventional approach.[9,10] As examples, contemporary clinical glaucoma trials, such as the Ocular Hypertension Treatment Study (OHTS),[11,12] the Low-Pressure Glaucoma Treatment Study (LoGTS)[13] and the European Glaucoma Prevention Study (EGPS),[14,15] used human observers to grade conventional stereo photos to evaluate the optic disc.

We are adapting modern computer vision methods to optic disc diagnosis in glaucoma but were concerned with the limited number of published comparisons of digital and film imaging formats for optic disc photographs, as ophthalmic photography has migrated from film-based to digital images. The few published reports mostly have assessed the utility of digital optic nerve images for quantitative parameters long used in glaucoma evaluation, such as the cup-to-disc ratio.[16–21] In contrast, in a considerably more extensive body of literature, the utility of digital images has been compared to that of film images in retinal diseases such as diabetic retinopathy[22–26] or macular degeneration.[27–29]

To address the suitability of digital stereo optic disc images as a platform for developing novel computerized approaches for glaucoma diagnosis, we assessed the quality and definition of optic disc features by directly comparing optic disc images obtained from individual subjects and displayed in different formats. We sought to avoid the well-known inaccuracies in trying to assign quantitative estimates such as the cup-to-disc ratio.[9,10] Specifically, we obtained stereo color optic disc images contemporaneously in both digital and 35-mm slide film formats in subjects with a range of optic disc cup sizes and included both scanned and grayscale images in the analysis.

Materials and Methods

Subjects

The optic nerves of both eyes of 34 outpatients at the Scheie Eye Institute were photographed. Images from five subjects were excluded: three did not have film photographs because subjective photophobia or poor fixation or cooperation prevented completion of the full photographic series; one had markedly blurred fundus images bilaterally from cataracts; and one had severe bilateral peripapillary choroidal neovascularization and atrophy from presumed ocular histoplasmosis that was too advanced to permit meaningful optic disc assessments. Two observers, (DJP and RAS) different from those who subsequently graded the images, selected the digital and film images of one eye from the remaining 29 subjects, using subjective assessment of the image qualities. The images were obtained from subjects aged 27 to 87 years, 16 (55%) of 29 of whom were women, of various ethnicities (16 [55%] Caucasian, 12 [41%] African American, and 1 [3%] East Indian). Since image qualities per se were the subject of this study, subjects were recruited to include optic discs with a broad range of cup-to-disc ratios but without regard to visual field status or specific glaucoma diagnosis. Of these 29 subjects, 5 had a clinical diagnosis of glaucoma; 9 had suspected glaucoma diagnosed on the basis of either intraocular pressure or optic disc asymmetry; and 15 had a diagnosis unrelated to glaucoma. All subjects had undergone an ophthalmic examination by their staff physicians, who had determined that they required pupil dilation as part of that day's examination. The subjects provided informed consent before pupil dilation, which was performed with topical drops according to their physician's instructions. The University of Pennsylvania Institutional Review Board approved this study, and the research was conducted in accordance with the Declaration of Helsinki for research involving human subjects.

Photography

After pharmacologic mydriasis, sequential stereo optic disc images were obtained at 1× magnification in both 35-mm slide film and digital formats by interchanging the camera back on a fundus camera (model FF4; Carl Zeiss Meditec, Inc., Dubin, CA). The 35-mm film images were captured on slide film (Fujichrome Velvia 100; Fujifilm USA, Inc., Valhalla, NY; ISO 100; http://www.fujifilm.com/products/professional_films/pdf/velvia_100_datasheet.pdf). The digital images were captured with a digital imaging system (OIS WinStation 5000; Ophthalmic Imaging Systems, Sacramento, CA) with a 4.9-megapixel Bayer sensor camera back (2392 × 2048 pixels; MegaVision; Santa Barbara, CA).

Image Visualization

Digital Image Display. Digital images were displayed on a 26-in. display monitor (MultiSync LCD2690WUXi; NEC Corp.; Tokyo, Japan) set to a screen resolution of 1920 × 1200 pixels, a native RGB setting of 6500°K, and brightness and contrast settings of 50.3% and 50.0%, respectively. The monitors were calibrated with a colorimeter (i1Display 2; X-Rite Inc., Grand Rapids, MI) before use each day, with white point values in the range of 6400 to 6700°K. The digital stereo images were viewed through a stereoscope (Screen-Vu; PS Mfg, Portland, OR).

35-mm Film Images. Slide film images were viewed on a horizontally placed fluorescent light box with 5000°K color-corrected lamp (Porta-Trace; Gagne, Inc., Johnson City, NY) with a Donaldson stereo viewer (GJ Davco, Holbrook, MA).

Images

From either the primary digital images or the film images, two investigators (DJP and RAS) selected the optimum stereo image pair for each subject for use throughout the investigation, based on subjective assessment of image clarity, stereo effect, minimal artifacts, and overall image quality.

35-mm Slide Film Digitalization. For each subject, a single pair of stereoscopic 35-mm slide film images was scanned at 250 dpi resolution (Super Coolscan 5000 ED; Nikon Corp.; Tokyo, Japan) using scanning software (SilverFast Ai Studio; ver. 6.5.5r5 LaserSoft Imaging AG; Kiel, Germany) that gave the scanned images the approximate dimension of the primary digital images (9.6 inches horizontally × 8.2 inches vertically), with a Q Factor (a setting for adjusting scan resolution) of 1.0. These scan settings yielded images of size and pixel dimensions approximately identical with those of the primary OIS images. Because of the camera's Bayer sensor, digital OIS WinStation images are automatically sharpened. To render the scanned images comparable visually to the primary digital images, the scanned film images were converted from RGB to LAB mode (Photoshop CS3 Extended; Adobe Systems Inc.; San Jose, CA), the lightness channel was highlighted, and an Unsharp Mask filter was applied with settings (amount 100%; radius 2.5 pixels; threshold 0) as recommended by OIS (B. Yates, OIS, personal communication, 2008). The image was then converted back to RGB mode and saved in TIFF format.

Grayscale Images. To convert the primary stereo digital color images to grayscale, two investigators (DJP and RAS) evaluated various conversion protocols on images with different cup-to-disc ratios for both the quality of the optic disc image, including its stereo, and the visibility of the peripapillary nerve fiber layer. From these assessments, grayscale images were created from the original color digital images (Channel Mixer adjustment layer, with the following proportions: red 0%; green 85%; blue 15%; Photoshop, Adobe Systems).

Digital Image Displays. For display on a computer monitor, all images were prepared at 25% magnification (using Photoshop; Adobe Systems). The fundus images were cropped with a fixed-size elliptical marquee tool, so that the optic disc was centered in a 1975-pixel-diameter circle. The stereo image pairs were mounted onto a digital canvas measuring 24 × 17 inches (horizontal by vertical; 6000 × 4250 pixels) with a black background.

To compare primary digital images to the 35-mm film images, we mounted the stereo digital image pair alone for digital display.

The primary versus scanned digital images were compared by displaying one stereo pair of primary digital images and one stereo pair of scanned 35-mm film images for each subject, one above the other, positioned in the top or bottom location in random order; graders were masked to the location of image types on the display. For digital color versus grayscale comparisons, the primary digital stereo color images and their corresponding grayscale images were similarly mounted on a black canvas as stereo pairs, one above the other. Whether the stereo images were in color or grayscale could not be masked, but the color or grayscale images were placed in the top or bottom display position in random order in case monitor location might influence grading.

Primary Image Format Comparisons and Grading Criteria

To obtain a broad perspective, two experienced glaucoma subspecialists (EM, JP-S), the head of the Scheie Fundus Reading Center (JA), and an experienced ophthalmic clinical photographer (WN) each independently compared the primary digital images (images acquired using the digital camera back) to the three other image formats: (1) 35-mm slide images (images acquired during the same photography session using the 35-mm film camera back); (2) scanned film images (digital images obtained by scanning the 35-mm film images); and (3) grayscale images (images from converting primary color digital images to grayscale digital images).

Unless otherwise specified, the image formats were assessed for 19 parameters (listed on Tables 1–3) as being equivalent, one type being slightly superior or one type being very superior, resulting in a 5-point grading scale. If a particular parameter was judged as not present or indeterminate, it was excluded from subsequent analysis. Each grader answered every query. For classification purposes only, the graders used the images from the primary digital versus scanned film comparison to estimate the vertical and horizontal cup-to-disc ratios; the mean of the vertical and horizontal cup-to-disc ratios from all graders was used to assign the cup-to-disc ratio for each subject.

Secondary Image Format Comparisons

Three observers (EM, JP, and WN) conducted two sets of secondary image format evaluations: (1) intraobserver reproducibility, and (2) monoscopic versus stereo viewing. For both secondary comparisons, they graded the images only along the global parameter of "overall gestalt" (parameters 18 and 19 in Tables 1–3). The intraobserver reproducibility was assessed by having the observers evaluate the same image sets at a later date under conditions identical with those of the initial more detailed comparisons. Monoscopic digital images were the identical digital images viewed on a monitor without using a stereoscope; the 35-mm slide images were evaluated monocularly with a single lens of the same stereo viewer.

Data Analysis

Image Format Preference Comparisons. Because few gradings ranked one image format as "very superior" to the other, the gradings were compressed into a 3-point scale (i.e., formats equivalent, one format superior, or the other format superior). To compare the gradings of the primary digital images to each of the other three image formats, we developed a preference ratio R calculated according to $R = (\text{DIG} + \text{E})/(\text{ALT} + \text{E})$, where DIG is the number of responses grading the primary digital image format as superior, E is the number of responses judging the two image formats as equivalent, and ALT is the number of responses grading the alternative image format as superior. This preference ratio R accounted for both the equivalent and preference ratings such that the analysis of equivalent ratings also influenced the outcome (e.g., higher proportions of equivalent ratings correspond to R closer to 1). An R was calculated for every query in each of the image format comparisons. Thus, $R = 1$ indicated that the readers overall judged the two image formats as equivalent for a particular parameter; $R > 1$ indicated a preference among the readers for the primary digital image format; $R < 1$ indicated a preference for the alternative image format.

To decide whether R was significantly different from 1 (i.e., the proportion of one image format judged superior being different from the proportion of the other format judged superior), the comparison of the superior proportion between two image formats was performed by using the Generalized Estimating Equation (GEE) approach,[30] to account for the grading of the same images by multiple graders. For this analysis, gradings of "equivalent" were excluded because they do not contribute any statistical information on superiority (analysis executed through PROC GENMOD in SAS ver. 9.1; SAS Inc., Cary, NC). The dependent variable was the preference of image format (coded as 1 for the primary digital image superior, and 0 for alternative image format superior), and the gradings from the same image sets were identified by an image ID number. The binomial distribution was specified; and an independent working correlation structure was used. In this analysis, larger deviations of R from 1 do not necessarily yield smaller P values, as the correlation among gradings by multiple graders may differ for different image parameters, leading to different variance estimates and resulting P values. The Bonferroni correction for multiple comparisons was used with 19 parameters for each set of the image format comparisons in Tables 1 to 3, and statistical significance was therefore defined as $P \leq 0.0026$ (i.e., 0.05/19). The κ statistics, weighted κ, and percentage of agreement were calculated to assess the concurrence of the observers for the image assessments and to assess intraobserver agreement. For these calculations, the gradings were first compressed into a three-level scale (i.e., one-image format superior to the comparator format, formats equivalent, or the comparator format superior). In calculating the weighted κ statistic, Cicchetti-Allison weights[31] were calculated by using the formula: $W_{ij} = 1 - [\text{absolute value } (C_i - C_j)/(C_3 - C_1)]$, where W_{ij} is the weight for the cell in column i and row j in a 3×3 contingency table, and C_i is the value of the index for column or row i. The formula yields the following weights: $W_{11} = W_{22} = W_{33} = 1$, $W_{12} = W_{23} = 0.5$, and $W_{13} = 0$. In comparing monoscopic to stereoscopic viewing (see Table 5), only four comparisons were made for each of the three image format comparisons, and the level of statistical significance was defined as $P \leq 0.0125$ (i.e., 0.05/4) per the Bonferroni correction.

Cup-to-Disc Ratio. As the cup-to-disc ratio is a prominent feature for glaucoma evaluation, we stratified the images based on whether each subject's ratio was above or below the median ratio of 0.475. The preferences in the stratified analyses in each of the format assessments were evaluated with the GEE for parameters, using the Bonferroni correction for comparison of 38 parameters, as the 19 parameters were assessed for both large and small optic disc cups. Thus, statistical significance was achieved with $P \leq 0.0013$ (i.e., 0.05/38). Intraclass correlations were used to assess interobserver agreement in quantifying the cup-to-disc ratios.

Statistical Power for Image Format Preference Comparisons. Power calculations showed that the sample size of the study (29 sets of images graded by four graders) would provide very high power (84%–93%) to detect a 30% difference in preference percentage and moderate power (48%–62%) to detect a 20% difference in preference percentage when the correlations among gradings by four graders is low (correlation coefficient = 0.25). When the correlation is higher (correlation coefficient = 0.50), the statistical power becomes lower, ranging from 69% to 83% for detecting a 30% difference, and ranging from 35% to 48% for detecting a 20% difference in preference percentage.

Results

In comparing primary digital images to 35-mm slide film images (Table 1), the image qualities and ease of defining optic disc features was generally judged superior in the digital images. Although the assessment of optic cup diameter was equivalent between these two formats, there was an overall preference for primary digital images considering all parameters. One of the observers generally preferred 35-mm slide film images in the overall gradings, however, whereas the other three all preferred the primary digital format.

When the 35-mm slides were digitized and compared to the primary digital images taken at the same photography session (Table 2), the observers also judged some qualities of the primary digital images as superior. For defining features of the optic disc and assessing the optic cup diameters, though, the two formats were seen as equivalent. There was no preference in the overall gestalt comparing primary digital to scanned optic disc images.

With a few notable exceptions, color and grayscale optic disc images were judged as mostly equivalent (Table 3). Peripapillary atrophy was more easily defined in many of the color images. The quality of the nerve fiber layer was superior in the grayscale images, as was the ease of defining the paths of blood vessels on the optic disc.

As typical for subjective judgments of cup-to-disc ratios,[6,9] the ratings by the four observers for cup-to-disc ratio showed moderate intraclass correlation: horizontal orientation = 0.70 (95% confidence interval [CI], 0.56–0.83); vertical orientation = 0.65 (95% CI, 0.49–0.80); average of horizontal and vertical orientations = 0.69 (95% CI, 0.54–0.82). The median cup-to-disc ratio (0.475) was used only for stratifying the images into two groups.

Stratifying images as greater or lesser than the median cup-to-disc ratio of the four graders (0.475) had only minor impact on the grading. In the primary digital versus 35-mm slide comparisons, the observers judged digital images superior for assessing blood vessel paths in optic discs with large cup-to-disc ratios (parameter 12, Table 1; $R = 1.62$; $P = 0.001$), but they judged the two formats as equivalent for images with small cup-to-disc ratios, accounting for the overall lack of preference for this parameter (Table 1). Digital images also were judged superior to 35-mm slides for defining both the vertical cup diameter (parameter 13, Table 1; $R = 1.39$; $P = 0.0004$) and horizontal cup diameter (parameter 14, Table 1; $R = 1.39$; $P = 0.001$) in optic discs with large cups, thus accounting for the overall preference for digital images in judging cup diameters (Table 1). In comparing the primary digital images to scanned 35-mm slide images, blood vessel edges (parameter 2 in Table 2; $R = 1.96$; $P < 0.0001$) and the

Table 1. Preferred Image Formats For Stereo Optic Disc Images: Primary Digital vs. 35-mm Slide Film Images

No.	Parameter	Preferred Format	Observer Preferences (%)†			R‡
			Primary Digital	Equivalent	35-mm Slide Film	
Overall Image Quality						
1	Image clarity, overall	Primary digital	59.5	19.8	20.7	1.96***
2	Definition of blood vessel edges (where in best focus)	Primary digital	39.7	44.8	15.5	1.40**
3	Color quality	Primary digital	50.9	31.0	18.1	1.67***
4	Image smoothness (absence of grain)	Primary digital	36.2	59.5	4.3	1.50***
5	Stereo quality in image sets	Primary digital	56.9	33.6	9.5	2.10***
6	Overall quality of nerve fiber layer	Primary digital	48.5	35.0	16.5	1.62**
Ease of Defining Optic Disc Features						
7	Optic disc edge	Primary digital	26.1	70.4	3.5	1.31***
8	Peripapillary atrophy	Primary digital	33.3	63.4	3.2	1.45***
9	Neuroretinal rim	Primary digital	26.7	62.1	11.2	1.21**
10	Optic cup margin	Equivalent	27.6	57.8	14.7	1.18NS
11	Optic cup slope	Primary digital	36.2	53.4	10.3	1.41***
12	Blood vessel paths on optic disc, for defining the optic cup	Equivalent	38.8	39.7	21.6	1.28NS
13	Vertical optic cup diameter	Primary digital	26.7	61.2	12.1	1.20*
14	Horizontal optic cup diameter	Primary digital	26.7	61.2	12.1	1.20*
15	Optic cup depth	Primary digital	50.0	37.9	12.1	1.76***
Optic Cup Diameters: Comparison of Size between Images (rating longer apparent diameter)						
16	Vertical optic cup diameter	Equivalent	7.8	85.3	6.9	1.01NS
17	Horizontal optic cup diameter	Equivalent	7.8	84.5	7.8	1.00NS
Overall Gestalt						
18	Overall gestalt of optic disc images	Primary digital	64.7	17.2	18.1	2.32***
19	Including forced choice if initial grading was equivalent	Primary digital	74.1	0	25.9	2.87***

† % of ratings judging the indicated format as superior/very superior or equivalent to the other format.

‡ R is the preference ratio = (DIG + E)/(ALT + E), where DIG is the % of responses grading the primary digital image format as superior, E is the % of responses judging the two image formats as equivalent, and ALT is the % of responses grading the 35-mm slides (i.e., the alternative image set) as superior.

* $P \leq 0.0026$; ** $P \leq 0.001$; *** $P \leq 0.0001$ (Bonferroni corrected), for $R \neq 1$. NS, R not significantly different from 1 (i.e., equivalence of image formats).

overall quality of the retinal nerve fiber layer (NFL; parameter 6 in Table 2; $R = 2.10$; $P = 0.0012$) were each judged superior in the primary digital images in optic discs with small cup-to-disc ratios; but the formats were equivalent in optic discs with large cup-to-disc ratios. In comparing primary color digital images to digital grayscale images, the ease of defining the blood vessel paths on the optic disc was judged superior in grayscale images only for optic discs with small cup-to-disc ratios (parameter 12, Table 3; $R = 0.80$; $P = 0.0005$). No other parameters were differentially affected by the size of the cup-to-disc ratio.

Interobserver and Intraobserver Agreement

In general, there was considerable interobserver variability in the preferred format for the different parameters in the optic discs evaluated here, confirmed statistically with low κ values (range, −0.18 to 0.41; see Supplementary Table S1, http://www.iovs.org/cgi/content/full/51/11/5667/DC1). A consistent difference between observers occurred in the 35-mm slide film comparison, where three observers preferred the digital format and one preferred film.

In the secondary image format comparisons, the three observers showed moderate agreement with their initial assessments when they compared the same image sets at a later time (Table 4). The percentage of agreement ranged from 60.5% to 83.9%, depending on the image format. The κ values also indicated moderate agreement for the repeat comparisons of the primary digital images to either 35-mm slides or scanned images of the slides. In the comparison of the primary color digital images to the grayscale conversions, the κ values indicated only slight to fair agreement between the observers' initial and subsequent assessment of the overall gestalt parameter.

Monoscopic Viewing

The format preferences were mostly similar in the comparisons of monoscopic to stereoscopic viewing (Table 5). The observers preferred the primary digital images over the 35-mm slide film images in either monoscopic or stereoscopic viewing. The primary digital images and the scanned 35-mm slide film images were judged equivalent regardless of whether they were viewed in mono or stereo. On the basis of the forced-choice criterion, the color images were judged superior to grayscale images in the overall gestalt when viewed in either mono or stereo.

Discussion

This study demonstrates that primary or scanned (i.e., digitized) stereo optic disc images are reasonable substitutes for conventional 35-mm slide film images as a reference standard for the optic disc in glaucoma diagnosis. The findings also suggest that digital stereo optic disc images can serve as a productive platform for future development of novel computerized analytical approaches to optic disc analysis. The optic disc images in this study were selected to show a broad range of cup-to-disc ratios, an essential feature related to glaucoma

TABLE 2. Preferred Image Formats for Stereo Optic Disc Images: Primary Digital vs. Scanned 35-mm Slide Film Images

			Observer Preferences (%)†			
No.	**Parameter**	**Preferred Format**	**Primary Digital**	**Equivalent**	**Scanned Slides**	***R*‡**
Overall Image Quality						
1	Image clarity, overall	Equivalent	43.1	29.3	27.6	1.27^NS^
2	Definition of blood vessel edges (where in best focus)	Primary digital	50.0	31.0	19.0	1.62*
3	Color quality	Primary digital	53.0	20.0	27.0	1.56**
4	Image smoothness (absence of grain)	Primary digital	21.6	73.3	5.2	1.21*
5	Stereo quality in image sets	Equivalent	33.6	42.2	24.1	1.14^NS^
6	Overall quality of nerve fiber layer	Primary digital	56.2	28.1	15.7	1.92*
Ease of Defining Optic Disc Features						
7	Optic disc edge	Equivalent	34.5	52.6	12.9	1.33^NS^
8	Peripapillary atrophy	Equivalent	38.2	47.4	14.5	1.38^NS^
9	Neuroretinal rim	Equivalent	27.6	47.4	25.0	1.04^NS^
10	Optic cup margin	Equivalent	30.2	43.1	26.7	1.05^NS^
11	Optic cup slope	Equivalent	32.8	40.5	26.7	1.09^NS^
12	Blood vessel paths on optic disc, for defining the optic cup	Equivalent	30.2	44.0	25.9	1.06^NS^
13	Vertical optic cup diameter	Equivalent	20.7	62.9	16.4	1.05^NS^
14	Horizontal optic cup diameter	Equivalent	21.6	59.5	19.0	1.03^NS^
15	Optic cup depth	Equivalent	27.6	48.3	24.1	1.05^NS^
Optic Cup Diameters: Comparison of Size between Images (rating longer apparent diameter)						
16	Vertical optic cup diameter	Equivalent	14.7	73.3	12.1	1.03^NS^
17	Horizontal optic cup diameter	Equivalent	17.2	70.7	12.1	1.06^NS^
Overall Gestalt						
18	Overall gestalt of optic disc images	Equivalent	46.6	24.1	29.3	1.32^NS^
19	Including forced choice if initial grading was equivalent	Equivalent	56.9	0	43.1	1.32^NS^

Data and comparisons are as explained in the footnote to Table 1.

diagnosis. Of importance, in the primary comparisons in this study, we used stereoscopic, not monoscopic, optic disc images, and the secondary comparison of monoscopic to stereoscopic images suggested that observer overall preferences for image formats would change little with monoscopic viewing. As with many other approaches to subjective evaluation of the optic disc,[9,10] the intraobserver reproducibility of the gradings was moderate.

In comparison to 35-mm slides acquired at the same photographic session, three of four observers judged the digitally acquired stereo optic disc images viewed on a monitor as superior for most parameters important in glaucoma assessments (Table 1). These comparisons necessarily are dependent on characteristics of both the film and the digital system. We assessed only one film type, Fuji Velvia 100 slide film, but Fujichrome films have been used for optic disc grading in other glaucoma research.[32] Since the scanned images subjectively reproduced the color characteristics of the 35-mm slides, the more equivalent ratings in the comparison of the primary digital to scanned slide images suggest that imaging qualities of the 35-mm slide film per se do not fully account for the preferences for viewing images in the digital format.

Features of our digital imaging system and differences in image presentation likely accounted for the observers' judgments. Whereas the slides were viewed on a horizontal light box, the digital images were displayed in a larger format on a calibrated high-resolution, broad-gamut, large-screen monitor. Although the resolution of the digital camera used in the study falls below that of some recently introduced fundus cameras, its 4.9-megapixel resolution conforms to what is thought to be an acceptable substitute for 35-mm slide film in fundus photography.[33] Even though the resolution of TIFF images is much greater than that of our monitor, we purposely avoided compressed image formats and their potential for distortion of image features because of our intent to use the current results to support our future plan for computational approaches to digital optic disc images. We recognize that certain features of this approach to digital image display may not be practical for non-research clinical environments.

Scanned images of the same 35-mm slides were judged as more equivalent to primary digital images of the optic discs (Table 2) than were the 35-mm slides to the same primary digital images. The image smoothness (i.e., absence of grain) was judged better in the primary digital images, suggesting that the digital camera used in this study is neither inducing significant electronic noise nor meaningfully sacrificing resolution compared with 35-mm slides. While blood vessel edge definition, color quality, and the nerve fiber layer also were judged superior in the primary digital images, the other parameters were considered equivalent. It is speculative why the observers considered the scanned images more comparable to the primary digital images than the 35-mm slides from which the scanned images were made. The scanned and primary digital images were presented together at equivalent magnifications on a black digital canvas on a monitor, a presentation probably easier for subjective comparisons than the need to switch to a light box for viewing 35-mm slides.

One other comparison was included: primary-color digital images versus digital grayscale conversions of the same images. In computer vision analyses of stereo images with a restricted color range, like fundus images, color does not generally provide additional information over luminance in grayscale images for disparity mapping and hence for stereo reconstruction. Accordingly, we hypothesized that grayscale would actually be a useful alternative to color for subjective assessments of optic disc images by providing disparity (i.e., stereo) and luminance information without distractions from color.

TABLE 3. Preferred Image Formats for Stereo Optic Disc Images: Primary Color Digital vs. Digital Grayscale Images

No.	Parameter	Preferred Format	Observer Preferences (%)† Color Digital	Equivalent	Grayscale Digital	*R*‡
Overall Image Quality						
1	Image clarity, overall	Equivalent	15.5	73.3	11.2	1.05^{NS}
2	Definition of blood vessel edges (where in best focus)	Equivalent	10.3	69.8	19.8	0.89^{NS}
3	Tonal quality	Equivalent	12.9	81.0	6.0	1.08^{NS}
4	Image smoothness (absence of grain)	Equivalent	4.3	94.8	0.9	1.04^{NS}
5	Stereo quality in image sets	Equivalent	3.4	87.9	8.6	0.95^{NS}
6	Overall quality of nerve fiber layer	Grayscale	8.7	38.5	52.9	0.52***
Ease of Defining Optic Disc Features						
7	Optic disc edge	Equivalent	19.0	74.1	6.9	1.15^{NS}
8	Peripapillary atrophy	Color	29.8	64.9	5.3	1.35***
9	Neuroretinal rim	Equivalent	12.9	82.8	4.3	1.10^{NS}
10	Optic cup margin	Equivalent	9.5	80.2	10.3	0.99^{NS}
11	Optic cup slope	Equivalent	8.6	79.3	12.1	0.96^{NS}
12	Blood vessel paths on optic disc, for defining the optic cup	Grayscale	3.4	77.6	19.0	0.84**
13	Vertical optic cup diameter	Equivalent	4.3	89.7	6.0	0.98^{NS}
14	Horizontal optic cup diameter	Equivalent	4.3	89.6	6.1	0.98^{NS}
15	Optic cup depth	Equivalent	3.4	87.9	8.6	0.95^{NS}
Optic Cup Diameters: Comparison of Size between Images (rating longer apparent diameter)						
16	Vertical optic cup diameter	Equivalent	1.7	94.0	4.3	0.97^{NS}
17	Horizontal optic cup diameter	Equivalent	1.7	96.6	1.7	1.00^{NS}
Overall Gestalt						
18	Overall gestalt of optic disc images	Equivalent	22.4	65.5	12.1	1.13^{NS}
19	Including forced choice if initial grading was equivalent	Equivalent	61.7	0	38.3	1.61^{NS}

Data and comparisons are as explained in the footnote to Table 1.

To convert color to grayscale images, our approach was guided by long-known chromatic properties of fundus images,[34,35] as pertinent to optic disc and nerve fiber assessments in glaucoma.[9] Red-free (i.e., green) light is absorbed well by hemoglobin and hence is useful to enhance visualization of blood vessels, a valuable parameter for assessing the stereo properties of optic discs. Although blue light does not materially contribute to the fundus image, imaging the fundus in blue light enhances the visualization of the nerve fiber layer, perhaps because it penetrates the retina poorly and is reflected superficially. Consistent with these concepts, we found in initial assessments that the green and blue channels mixed in an 85:15 proportion in Photoshop provided a grayscale conversion with excellent stereo and apparently improved nerve fiber visualization.

In comparing grayscale and color images, color added little to the quality of optic disc parameters, and grayscale was judged superior to color for assessing blood vessels and the nerve fiber layer (Table 3), two important parameters for glaucoma diagnosis. In the only exception to this generalization, color images were judged potentially superior for assessing peripapillary atrophy. Photoshop allows many methods to convert color to grayscale, and it is possible that a conversion technique not assessed here would have provided even more favorable grayscale stereo disc images. We are aware of no prior direct comparisons of color versus

TABLE 4. Intraobserver (i.e., Grade/Regrade) Agreement, for Three Observers

Parameter	Percentage Agreement	Weighted κ (95% CI)*
Primary Digital vs. 35-mm Slide Film Image Comparison		
Overall gestalt of optic disc images	70.1	0.55 (0.39–0.72)
Including forced choice if initial grading was equivalent	83.9	0.63 (0.45–0.80)
Primary Digital vs. Scanned 35-mm Slide Film Image Comparison		
Overall gestalt of optic disc images	60.9	0.52 (0.37–0.66)
Including forced choice if initial grading was equivalent	81.6	0.63 (0.47–0.79)
Primary Color Digital vs. Grayscale Digital Image Comparison		
Overall gestalt of optic disc images	72.4	0.38 (0.17–0.58)
Including forced choice if initial grading was equivalent	60.5	0.15 (−0.06–0.36)

* For the forced-choice parameter, κ values are shown because there were only two possible choices and the weighted κ and κ are thus equivalent.

Table 5. Monoscopic vs. Stereoscopic Image Gradings, for Three Observers

Primary Digital vs. 35-mm Slide Film Image Comparison

Parameter	Image Set	Preferred Format	Observer Preferences (%)† Primary Digital	Equivalent	35-mm Slide Film	R‡
Overall gestalt of optic disc images	Mono	Primary digital	51.7	24.1	24.1	1.57*
Overall gestalt of optic disc images	Stereo	Primary digital	54.0	24.1	21.8	1.70***
Including forced choice if initial grading was equivalent	Mono	Primary digital	65.5	0	34.5	1.90**
Including forced choice if initial grading was equivalent	Stereo	Primary digital	71.3	0	28.7	2.48***

Primary Digital vs. Scanned 35-mm Slide Film Image Comparison

Parameter	Image Set	Preferred Format	Observer Preferences (%)‡ Primary Digital	Equivalent	Scanned Slides	R‡
Overall gestalt of optic disc images	Mono	Equivalent	48.3	19.5	32.2	1.31NS
Overall gestalt of optic disc images	Stereo	Equivalent	31.0	37.9	31.0	1.00NS
Including forced choice if initial grading was equivalent	Mono	Equivalent	58.6	0	41.4	1.42NS
Including forced choice if initial grading was equivalent	Stereo	Equivalent	56.3	0	43.7	1.29NS

Primary Color Digital vs. Grayscale Digital Image Comparison

Parameter	Image Set	Preferred Format	Observer Preferences (%)† Color Digital	Equivalent	Grayscale Digital	R‡
Overall gestalt of optic disc images	Mono	Color	36.8	50.6	12.6	1.38*
Overall gestalt of optic disc images	Stereo	Equivalent	16.1	79.3	4.6	1.14NS
Including forced choice if initial grading was equivalent	Mono	Color	67.8	0	32.2	2.11*
Including forced choice if initial grading was equivalent	Stereo	Color	68.6	0	31.4	2.18*

The stereo image set data represent the replicate gradings by the three observers who conducted the monoscopic image assessments. Data and comparisons are as explained in the footnote to Table 1.

grayscale images in the evaluation of optic disc images, particularly for stereo, and our results suggest that the post-acquisition conversion of color to grayscale digital images would prove advantageous for at least some fundus features pertinent to glaucoma diagnosis.

In summary, primary digital images overall were rated equivalent or superior to 35-mm slide images or to the scanned images of these slides by most observers and for most parameters. Thus, digital stereo optic disc images are useful for evaluating the optic disc and, like film, can serve as a reference standard for other approaches to glaucoma diagnosis, especially approaches based on stereo. Given the similar assessments of digital and film formats, digital optic disc images also can serve as a platform for advanced computer vision applications addressing glaucoma diagnosis.

References

1. DeLeón-Ortega JE, Arthur SN, McGwin G Jr, Xie A, Monheit BE, Girkin CA. Discrimination between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective optic nerve head assessment. *Invest Ophthalmol Vis Sci.* 2006;47:3374–3380.
2. Reus NJ, de Graff M, Lemij HG. Accuracy of GDx VCC, HRT I, and clinical assessment of stereoscopic optic nerve head photographs for diagnosing glaucoma. *Br J Ophthalmol.* 2007;91:313–318.
3. Badalà F, Nouri-Mahdavi K, Raoof DA, Leeprechanon N, Law SK, Caprioli J. Optic disk and nerve fiber layer imaging to detect glaucoma. *Am J Ophthalmol.* 2007;144:724–732.
4. Pablo LE, Ferreras A, Fogagnolo P, Figus M, Pajarin AB. Optic nerve head changes in early glaucoma: a comparison between stereophotography and Heidelberg retina tomography. *Eye.* 2010;24:123–130.
5. Saito H, Tsutsumi T, Iwase A, Tomidokoro A, Araie M. Correlation of disc morphology quantified on stereophotographs to results by Heidelberg Retina Tomograph II, GDx Variable Corneal Compensation, and visual field tests. *Ophthalmology.* 2010;117:282–289.
6. Caprioli J, Jonas J, Vasile C. Optic disc photographs. In: Weinreb RN, Greve EL, eds. *Glaucoma Diagnosis: Structure and Function*. The Hague, The Netherlands: Kugler Publications; 2004:39–46.
7. Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Use of progressive glaucomatous optic disk change as the reference standard for evaluation of diagnostic tests in glaucoma. *Am J Ophthalmol.* 2005;139:1010–1018.
8. Hitchings RA. Glaucoma: an area of darkness. *Eye.* 2009;23:1764–1774.
9. Airaksinen PJ, Tuulonen A, Werner EB. Clinical evaluation of the optic disc and retinal nerve fiber layer. In: Ritch R, Shields MB, Krupin T, eds. *The Glaucomas*. St. Louis: Mosby; 1996:617–657.
10. Jampel HD, Friedman D, Quigley H, et al. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol.* 2009;147:39–44.
11. Gordon MO, Kass MA, OHTS Study Group. The ocular hypertension treatment study: design and baseline description of participants. *Arch Ophthalmol.* 1999;117:573–583.
12. Kass MA, Heuer DK, Higginbotham EJ, et al. The ocular hypertension treatment study: a randomized trial determines that topical

ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol.* 2002;120:701–713.

13. Krupin T, Liebmann JM, Greenfield DS, et al. The low-pressure glaucoma treatment study (LoGTS): Study design and baseline characteristics of enrolled patients. *Ophthalmology.* 2005;112:376–385.
14. The European Glaucoma Prevention Study (EGPS) Group. Reproducibility of evaluation of optic disc change for glaucoma with stereo optic disc photographs. *Ophthalmology.* 2003;110:340–344.
15. The European Glaucoma Prevention Study (EGPS) Group. Results of the European glaucoma prevention study. *Ophthalmology.* 2005;112:366–375.
16. Shuttleworth GN, Khong CH, Diamond JP. A new digital optic disc stereo camera: intraobserver and interobserver repeatability of optic disc measurements. *Br J Ophthalmol.* 2000;84:403–407.
17. 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.
18. Morgan JE, Sheen NJL, North RV, Choong Y, Ansari E. Digital imaging of the optic nerve head: monoscopic and stereoscopic analysis. *Br J Ophthalmol.* 2005;89:879–884.
19. Khouri AS, Szirth G, Realini T, Fechtner RD. Comparison of digital and film stereo photography of the optic nerve in the evaluation of patients with glaucoma. *Telemed J E Health.* 2006;12:632–638.
20. Khouri AS, Szirth BC, Shahid KS, Fechtner RD. Software-assisted optic nerve assessment for glaucoma tele-screening. *Telemed J E Health.* 2008;14:261–265.
21. Xu J, Ishikawa H, Wollstein G, et al. Automated assessment of the optic nerve head on stereo disc photographs. *Invest Ophthalmol Vis Sci.* 2008;49:2512–2517.
22. Lim JI, LaBree L, Nichols T, Cardenas I. A comparison of digital nonmydriatic fundus imaging with standard 35-millimeter slides for diabetic retinopathy. *Ophthalmology.* 2000;107:866–870.
23. Tennant MTS, Greve MDJ, Rudnisky CJ, Hillson TR, Hinz BJ. Identification of diabetic retinopathy by stereoscopic digital imaging via teleophthalmology: a comparison to slide film. *Can J Ophthalmol.* 2001;36:187–196.
24. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM, The Digital Diabetic Screening Group. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol.* 2002;134:204–213.
25. Hansen AB, Sander B, Larsen M, et al. Screening for diabetic retinopathy using a digital non-mydriatic camera compared with standard 35-mm stereo colour transparencies. *Acta Ophthalmol Scand.* 2004;82:656–665.
26. Rudnisky CJ, Tennant MTS, Weis E, Ting A, Hinz BJ, Greve MDJ. Web-based grading of compressed stereoscopic digital photography versus standard slide film photography for the diagnosis of diabetic retinopathy. *Ophthalmology.* 2007;114:1748–1754.
27. van Leeuwen R, Chakravarthy U, Vingerling JR, et al. Grading of age-related maculopathy for epidemiological studies: is digital imaging as good as 35-mm film? *Ophthalmology.* 2003;110:1540–1544.
28. Scholl HPN, Dandekar SS, Peto T, et al. What is lost by digitizing stereoscopic fundus color slides for macular grading in age-related maculopathy and degeneration? *Ophthalmology.* 2004;111:125–132.
29. Somani R, Tennant M, Rudnisky C, et al. Comparison of stereoscopic digital imaging and slide film photography in the identification of macular degeneration. *Can J Ophthalmol.* 2005;40:293–302.
30. Zeger S, Liang K. Models for longitudinal data: a generalized estimating equation approach. *Biometrics.* 1988;44:1049–1060.
31. Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *Am J EEG Technol.* 1971;11:101–109.
32. Feuer WJ, Parrish RK II, Schiffman JC, et al. The ocular hypertension treatment study: reproducibility of cup/disk ratio measurements over time at an optic disc reading center. *Am J Ophthalmol.* 2002;133:19–28.
33. Hubbard LD, Danis RP, Neider MW, et al. Brightness, contrast, and color balance of digital versus film retinal images in the Age-Related Eye Disease Study 2. *Invest Ophthalmol Vis Sci.* 2008;49:3269–3282.
34. Behrendt T, Duane TD. Investigation of fundus ocular spectral reflectance photography. *Arch Ophthalmol.* 1966;75:375–379.
35. Delori FC, Gragoudas ES, Francisco R, Pruett RC. Monochromatic ophthalmoscopy and fundus photography: the normal fundus. *Arch Ophthalmol.* 1977;95:861–868.