Doctoral dissertation in medicine

Developing new automated matched alternation flicker using optic disc photography for the detection of glaucoma progression

Major in Medicine
Department of Medical Sciences

The Graduate School, Ajou University

Marvin Lee

Developing new automated matched alternation flicker using optic disc photography for the detection of glaucoma progression

지도교수 안 재 홍

이 논문을 의학 박사학위 논문으로 제출함.

2018 년 8 월

아 주 대 학 교 대 학 원

의 학 과

이 마 빈

이마빈의 의학 박사학위 논문을 인준함.

심사 위원장	양 홍 석	인
심 사 위 원	안 재 홍	인
심 사 위 원	이 은 소	인
심 사 위 원	국 경 훈	인
심 사 위 원	노 승 수	인

아 주 대 학 교 대 학 원

2018년 5월 30일

Acknowledgement

For my beloved wife, Eun Jung Lee

and

my most precious treasure, Haewoon Lee

- Abstract -

Purpose: To evaluate a progression-detecting algorithm for a new automated matched alternation flicker (AMAF) in glaucoma patients.

Methods: Open-angle glaucoma patients with a baseline mean deviation of visual field (VF) test > -6 dB were included in this longitudinal and retrospective study. Functional progression was detected by two VF progression criteria and structural progression by both AMAF and conventional comparison methods using optic disc and retinal nerve fiber layer (RNFL) photography. Progression-detecting performances of AMAF and the conventional method were evaluated by agreement between functional and structural progression criteria. RNFL thickness changes measured by optical coherence tomography (OCT) were compared between progressing and stable eyes determined by each method.

Forty-five pairs of fundus photographs of normal tension glaucoma patients were collected. Fundus photographs were taken at intervals of more than 12 months. Alternating flicker images were created using a new AMAF application. In a blinded manner, two glaucoma experts and two general ophthalmologists compared the presence of glaucomatous structural changes using either the AMAF method or the side-by-side comparison method. The interobserver and intraobserver agreements were compared using the Bland and Altman plot analysis.

Results: Among 103 eyes, 47 (45.6%), 21 (20.4%), and 32 (31.1%) eyes were

evaluated as glaucoma progression using AMAF, the conventional method, and guided progression analysis (GPA) of the VF test, respectively. The AMAF showed better agreement than the conventional method, using GPA of the VF test (= 0.337; P < 0.001 and = 0.124; P = 0.191, respectively). The rates of RNFL thickness decay using OCT were significantly different between the progressing and stable eyes when progression was determined by AMAF (-3.49 ± 2.86 m/year vs. -1.83 ± 3.22 m/year; P = 0.007) but not by the conventional method (-3.24 ± 2.42 m/year vs. -2.42 ± 3.33 m/year; P = 0.290).

In inter-reader agreement analysis, The glaucoma experts detected more glaucoma progression using the AMAF method (average, 40%) compared with the side-by-side method (average, 23.3%). General ophthalmologists detected more glaucomatous progression with the AMAF method (average, 28.9%) than with the side-by-side method (average, 18.9%). The AMAF method showed fair to substantial interreader agreement (k = 0.530~0.645) and fair to perfect intrareader agreement (k = 0.503~0.955). Interreader and intrareader agreement using the AMAF method was better for the glaucoma experts compared with the general ophthalmologists.

Conclusions: The AMAF was better than the conventional comparison method in discriminating structural changes during glaucoma progression, and showed a moderate agreement with functional progression criteria. Regarding inter- and intrareader agreement, agreement for the glaucoma experts was best using the

AMAF method.

Keywords : Automated matched alternation flicker, Fundus photography, Progression of glaucoma

Table of Contents

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Glaucoma is an irreversible optic neuropathy resulting from retinal ganglion cell death that leads to progressive visual field (VF) loss. Glaucomatous damage is characterized by structural changes of the optic nerve head (ONH) and retinal nerve fiber layer (RNFL), functional changes of the VF, and ultimate loss of central visual acuity. Structural changes may precede the development of the functional VF defects.[1, 2] Early detection of these changes around the ONH is therefore essential in diagnosing and monitoring glaucoma. In glaucoma clinics, optical coherence tomography (OCT) and ONH photography (with red-free RNFL photography) have been commonly used to detect structural deterioration associated with glaucoma.

New and more sensitive imaging devices to detect structural progression have been recently developed. Guided progression analysis (GPA) using Cirrus OCT (Carl Zeiss Meditec, Dublin, CA, USA) is an example of such a device that detects structural changes of RNFL using statistical methods.[3, 4] Careful analyses of sequential stereoscopic optic disc photographs have been widely used to determine structural changes of the ONH. To improve the detection ability of early ONH structural changes, a novel method using alternation flicker chronoscopy has been introduced. Alternating flickers can be achieved by manually aligning and

alternating images of fundus photographs taken sequentially with overlapping slide projectors or by manual registration of digital images.[5] This technique has been reported to be effective and comparable to photographic review in several studies.[6-8] Automated flicker has recently become available, rendering it possible to make automated flicker images using Matched-Flicker 8 software (ver. 1.2; EyeIC, Narberth, PA, USA). This approach has an excellent ability to detect ONH contour changes, parapapillary atrophy, and disc hemorrhages.[5, 9, 10]

We have developed a new algorithm, by collaborating with the engineering department, to make automated flickering images of ONH and RNFL photography for clinical purposes. To make matched alternation flicker images using two fundus photographs taken at different time points using different camera angles, each image needs to be processed to adjust the size and rotation angle. In addition, detection of changes in RNFL defects can be a problem if two fundus images have different brightnesses. An ideal algorithm should detect changes that are not associated with the image-capturing procedure, but with glaucoma progression itself. In the following report, we investigated the reproducibility of our algorithm for automatically matched alternation flicker (AMAF), and compared its ability to detect glaucoma progression with the conventional method using serial ONH and RNFL photographs. And after this investigation we conducted second study to compare the interobserver and intraobserver agreement rates of the traditional side-

by-side comparison and AMAF methods for the detection of glaucomatous structural changes. In this way, we want to investigate the reliability and usefulness of newly developed methods.

MATERIALS AND METHODS

Selection of patients

In this longitudinal and retrospective study, the medical records of patients who visited the glaucoma clinic from January 2010 to December 2012 at the Department of Ophthalmology, Ajou University Hospital, for glaucoma evaluation, were reviewed. The study adhered to the Declaration of Helsinki and was approved by the Institutional Review Board of Ajou University Hospital (AJIRB-MED-MDB-13-137).

Primary open-angle glaucoma (POAG) patients with baseline mean deviation (MD) values better than -6 dB were identified from the medical records. Baseline and last fundus photography examinations had an interval of ≥ 1 year. All patients received at least five VF tests for GPA using the Humphrey Field Analyzer (HFA) (Carl Zeiss Meditec). All patients underwent a comprehensive ophthalmic examination, including a medical history review, best-corrected visual acuity,

spherical equivalent of the refractive error, slit lamp biomicroscopy, central corneal thickness (DGH-500; DGH Technology, Exton, PA, USA), intraocular pressure (IOP) measurement using Goldman applanation tonometry, gonioscopy, and automated perimetry (HFA) using the Swedish interactive thresholding algorithm with standard 24–2 strategies. Stereo optic-disc photography (ODP) (AFC-210; NIDEK, Aichi, Japan) and time domain OCT (Stratus OCT; Carl Zeiss Meditec) were performed on the same day. Pupil dilation was not applied for all subjects, and all FPs were taken by the same examiner. VF tests, and ODP and OCT examinations, were performed within 3 months of other ocular examinations. All the patients in this study had an IOP < 22 mmHg at every visit with or without medication, and had gonioscopic open-angle glaucoma without secondary causes. The baseline VF tests showed at least two reproducible VF defects compatible with glaucoma according to Anderson's criteria for minimal abnormality in glaucoma.[11] RNFL defects, corresponding with VF defects, were noted in red-free RNFL photography and/or OCT.

Creating the application

This application was designed to help ophthalmologists detect differences between two fundus images taken at two different time points, by alternately

exposing the images. Due to the time gap between the two images, there are usually changes in scale, rotation, translation, brightness, and noise.

At first we made manually arranged flicker image. Two fundus images taken at different time points were superimposed on the same area to create a projected image. And by zooming in and out and rotating the late shot, we made sure that the two pictures were as close as possible. The resulting flicker image appeared to be superior to the parallel comparison method in determining the change. Of course, this method is too time consuming, and clinical application is impossible.

Next, we created an application that creates a flicker image semi-automatically. This method allows the program to automatically rotate, enlarge and reduce the image when the examiner views the two photographs and designates three identical geopolitical positions within the two different fundus photographs. This method also had time-consuming disadvantages, and serious distortions occurred when there was a mistake in specifying the point of the examiner, or when the photographing angle of the fundus photograph was greatly different.

A well-known computer vision technique, Speeded Up Robust Features (SURF),[12] was used to better match corresponding features of the two images. SURF facilitated feature correspondence that was scale- and rotation-invariant, and

also accounted for noise and photometric deformations. Using the feature correspondence, a transformation matrix was used to deform one image to another by solving a linear least squares optimization problem. The feature points in one image therefore had almost the same positions relative to their corresponding feature points as the other deformed image. Attempts to match all the parts in an image sometimes caused unwanted distortion of the image, especially around the ONH. To increase the degree of match around the ONH, we therefore added a function that allowed the examiners to specify an area as a reference for image fusion (Figure. 1). After this feature matching, the brightness levels of the two matched images were adjusted as closely as possible. This application provided users with controls for the scale, rotation invariants, and exposure frequency. The application was implemented using the C++ programming language, OpenCV, and Fast Light Toolkit libraries using the Windows 7 operating system (Microsoft Corp., Redmond, WA, USA).

Making flicker images and the determination of progression

Two original fundus photographs taken at two different time points for each patient were extracted from the fundus photography database. Using our flicker application for AMAF, the first and second images were selected and three

hertz flicker images were automatically created with contrast adjustments. If the examiner determined that the generated flicker images were sufficient to determine fundus changes, the glaucoma progression was evaluated by a single glaucoma specialist (ML) using a 3-point grading scale. Grade 3 was defined as definite glaucomatous change in the fundus images, and grade 2 was defined as a fundus change with suspected glaucoma. We only used the 3-point grading scale to evaluate glaucoma progression. If there was no detectable change in the flicker image, it was designated as grade 1. To compare the progression-detecting ability of AMAF with the parallel comparisons, glaucoma progression was evaluated in terms of four different ONH and RNFL changes, involving changes in the neuroretinal rim of the optic disc, in the path of the vessel around the optic disc, and in the widening and deepening of the RNFL defect (defined as a loss or decrease in RNFL striation leading to darkening of the previous defects).

Conventional side-by-side comparisons of fundus photographs were performed by the same examiner (ML) in a blinded manner. Another operator prepared two fundus photographs to make an automated flicker image with randomly assigned serial numbers. The inspector (ML) evaluated it in the same manner using the 3-point grading scale.

VF progression was determined independently after evaluating glaucoma

progression using AMAF and side-by-side comparisons in a blinded manner. VF progression was defined by GPA (integrated in the HFA) and by the collaborative initial glaucoma treatment study (CIGTS) criteria. A “likely progression” using the GPA software was defined as progression of GPA criterion. An increase in CIGTS score of three or more compared with an average of two baseline VF tests was defined as progression of CIGTS criterion.[13] A VF test conducted on the same day or within 3 months from the first fundus photography was used as a baseline VF. A VF test conducted at the closest time point (within 3 months) of the final fundus photography was used to determine glaucoma progression.

For progressing and stable eyes assessed by both AMAF and conventional ODP analyses, changes of average RNFL thickness using Stratus OCT were compared between the two methods.

Evaluation of agreement of AMAF for detection of glaucomatous structural change

Serial sets of fundus photographs of open-angle glaucoma (OAG) patients which were not included in previous investigation were re-selected in a consecutive manner from database of patients enrolled for the AMAF study. For producing

flicker images, two original FPs having time intervals at least 1 year apart were obtained from the AFC-210 database. Flicker images (4 Hz) were automatically prepared using the AMAF application. (Figure 2A). The progression of glaucoma was evaluated by two glaucoma experts (A and B) and two general ophthalmologists (C and D) using the “5-point grade” scale and the “yes or no” scale. In the 5-point scale, point 5 designated a definite glaucomatous change, point 4 designated a suspicious glaucomatous change, point 3 designated a change that was not related to glaucoma, point 2 designated a suspicious and non-glaucomatous change, and point 1 designated no change. In the ”yes or no” scale, “yes” designated a glaucomatous change, and “no” designated no changes found by the examiner.

As a contrast method to determine the agreement, the classical side-by-side comparison method of FPs was performed by the same examiners in a blinded manner (Figure 2B). Another operator prepared the same two fundus photographs used for the AMAF analysis with randomly assigned serial numbers, and the same inspectors evaluated the progression of glaucoma based on the same scales as the AMAF analyses. To assess interreader repeatability, two glaucoma experts and one of the general ophthalmologists conducted a second AMAF analysis and a second side-by-side comparison 1 week after the first session.

Statistical analysis

Descriptive results were expressed as means ± standard deviation, and *P* values < 0.05 were considered statistically significant. The kappa statistic was used to quantify and evaluate the agreement between evaluations of glaucoma progression determined by different methods. A kappa value of 0.1–0.2 was considered as slight agreement, 0.21–0.40 as moderate agreement, 0.41–0.60 as fair agreement, 0.61~0.80 as substantial agreement, and 0.81~0.99 as almost perfect agreement.[14] SPSS (ver. 20.0; SPSS Inc., Chicago, IL, USA) and MedCalc software (ver. 12.0; Mariakerke, Belgium) were used for analyses.

RESULTS

In total, 103 POAG patients were enrolled after screening 210 patients and excluding patients who did not meet the inclusion criteria. The mean age of the patients was 52.8 ± 9.9 years (range: 19–79 years), and 54 patients (52.4%) were male. The mean interval between the first baseline fundus photography and the final examination was 28.7 ± 9.0 months. The MD values decreased significantly from -0.93 ± 2.14 dB to -1.62 ± 2.93 dB, and both pattern standard deviation values and CIGTS scores significantly increased. RNFL thicknesses also decreased

significantly in all quadrants, including the average value (Table 1).

Kappa coefficients showed almost perfect intraobserver agreement (κ = 0.863, 95% CI = 0.806 ~ 0.919, P < 0.001) using AMAF analysis and substantial agreement (κ = 0.797, 95% CI = 0.716 ~ 0.878, P < 0.001) using conventional side-by-side comparisons of ODP.

We compared two methods for assessing structural progression based upon agreement with two progression criteria of functional tests using kappa statistics. The AMAF showed moderate agreement with both GPA (κ = 0.337) and CIGTS (κ= 0.241) progression criteria, whereas the conventional comparison method showed no significant agreement with two criteria for functional progression (Table 2). The AMAF and conventional comparison showed a fair agreement with each other (κ = 0.468, *P* <0.001).

The AMAF designated 47 of 103 eyes with glaucoma progression, whereas 21 eyes were designated with glaucoma progression using the conventional comparison method. The two methods showed a substantial agreement in detecting widening of RNFL defects (κ = 0.785; *P* < 0.001). RNFL defect depth changes (κ = 0.354; *P* < 0.001) and changes of the neuroretinal rim (κ = 0.241; *P* = 0.001) showed moderate agreement. Detection of glaucoma progression using a change in

path of the vessel around the ONH showed slight agreement (κ = 0.109; *P* = 0.015). AMAF detected changes of the vessel around the optic disc in 15 eyes but only a single eye was detected with a vessel change using the conventional comparison method (Table 3).

The rate of average RNFL thickness decrease of progressing eyes (-3.49 ± 2.86 μm/year) was significantly different from that of stable eyes (-1.83 ± 3.22 μm/year; *P* = 0.007) when progression was determined using AMAF. However, there was no significant different between progressing and stable eyes using the conventional method (-3.24 ± 2.42 μm/year vs. -2.42 ± 3.33 μm/year; *P* = 0.290). The average RNFL thickness of progressing eyes decreased significantly more than that of stable eyes when the GPA criterion was used (-3.81 ± 3.20 μm/year vs. -2.87 ± 3.86 μm/year; *P* = 0.008), whereas the rate of RNFL decrease between progressing and stable eyes was not significantly different when the CIGTS criterion was used (-3.63 ± 2.24 μm/year vs. -2.34 ± 3.31 μm/year; *P* = 0.101). When progression was determined by GPA combined with AMAF or the conventional method, the average RNFL thickness of progressing eyes decreased significantly more than that of stable eyes using both the AMAF + GPA criteria (-3.52 ± 3.02 μm/year vs. -1.47 ± 2.99 μm/year; *P* = 0.001) and the conventional comparison + GPA criteria. (-3.56 ± 2.98 μm/year vs. -1.86 ± 3.12 μm/year; *P* =

0.006) (Table 4).

Forty-five eyes from 45 OAG patients were used for evaluation of agreement of AMAF for detection of glaucomatous structural change in the final analyses. The mean age of the patients was 54.1 ± 9.8 years (range: 31~73 years), and 25 patients (56%) were male. In total, 35 eyes were right eyes and 10 eyes were left eyes. The mean interval between the baseline and the final examination was 676.6 ± 247.7 days. The mean deviation MD values decreased significantly from -2.92 ± 4.11 dB to -3.56 ± 4.45 dB (paired *t*-test, p = 0.033).

Table 5 shows the grading from each examiner. Using the 5-point scale, the glaucoma specialists (A and B) assessed the glaucomatous progression of 15 (A) and 21 (B) patients (average, 40%) using the AMAF method, and 7 (A) and 14 (B) patients (average, 23.3%) using the side-by-side comparison method. The general ophthalmologists (C and D) assessed the glaucomatous progression in 13 (C) and 13 (D) patients (average, 28.9%) using the AMAF method, and in 8 (C) and 9 (D) patients (average, 18.9%) using the side-by-side comparison method. Using a “yes or no” scale, the average values of glaucomatous changes in the patients were 51.1% and 40% using the AMAF method, as assessed by the glaucoma experts and the general ophthalmologists, respectively. Using the side-by-side comparison method, the average values of glaucomatous changes in the patients were 28.8% and 22.2%,

as assessed by the glaucoma experts and the general ophthalmologists, respectively (Table 5). The probability of finding glaucomatous changes increased when using the AMAF method for both examiner groups, but the increase in probability was greater in the glaucoma specialist group.

Kappa coefficients and ICC coefficients indicated that interreader agreement among the glaucoma specialist group was better using the AMAF method than the side-by-side comparison method, for both the 5-point scale and the “yes or no” scale (Table 6). However, the side-by-side comparison method showed better interreader agreement among the general ophthalmologists relative to the glaucoma specialists. The intrareader agreement of the glaucoma specialists was better using the AMAF method (kappa value = 0.910~0.955, ICC coefficient = 0.955~0.992) than the side-by-side comparison method (kappa value = 0.759~0.823, ICC coefficient = 0.863~0.965). In contrast, the intrareader agreement of the general ophthalmologists was worse using the AMAF method (kappa value = 0.503~0.548, ICC coefficient = 0.669~0.769) than the side-by-side comparison method (kappa value = 0.833~0.938, ICC coefficient = 0.884~0.969) (Table 6).

Inter-instrument agreement between the AMAF method and the side-by-side comparison method showed only a fair kappa coefficient(κ) value for the two glaucoma specialists and one of the general ophthalmologists (C) (Table 7).

DISCUSSION

Previous studies have reported that alteration flicker chronoscopy is more advantageous in detecting changes of the optic disc and peripapillary structures than conventional side-by-side comparison of fundus photographs.[5, 7, 15, 16] Automated matched flicker was developed using evolving imaging technology. It can better match the flicker images using simple operations, and has been reported to have good interreader and intrareader agreements.[8] It has also been reported that flicker chronoscopy was better than the conventional side-by-side comparison method for evaluating the progression of glaucoma, and was better at detecting changes in the contour of the ONH, detecting increases in the area of the peripapillary atrophy, and detecting the appearance and disappearance of disc hemorrhages.[5, 9, 10, 17]

Using AMAF to detect structural changes during glaucoma progression, we focused on changes in the shape of the neuroretinal rim of the ONH, changes of the vessel contour around the optic disc, widening of the RNFL defects, and deepening (defined as loss or decrease in RNFL striation leading to darkening of the previous defects even after adjusting brightness and contrast of images compared) of RNFL defects. Development of disc hemorrhages and peripapillary atrophy changes were not included in the progression criteria if there were no other

associated changes. Although the matching process is automated using AMAF, an examiner can directly specify the desired matching area using stable structural markers, including the optic disc margin and major blood vessels. The AMAF application can automatically adjust the brightness of the two fundus photographs to increase the brightness of the darker image before performing the automated matched flicker. These new features of AMAF improved the agreement between structural and functional progression criteria in our study (Table 2).

The agreement between structural and functional progression criteria and the ability to detect glaucoma progression are known to be influenced by both detection methods and the severity of glaucoma.[18, 19] In a recent observational cohort study, progression of glaucoma was detected by stereo photographs, GPA of the HFA, and GPA using the Cirrus spectral domain OCT in 6.9%, 15%, and 25.6% of eyes, respectively. In this study, normal, suspected glaucoma, and glaucoma patients were included in 246 eyes of 148 patients.[20] The detection rate of stereo photography was lower than that of the VF test in those patients with early stages of glaucoma. A higher detection rate of glaucoma progression was reported using flicker chronoscopy in another study, and 50 (48.5%) of 103 eyes showed at least one sign of structural progression, although the severity of glaucoma was not reported.[21] In our study, the conventional comparison method, AMAF, and GPA of the VF test could detect glaucoma progression in 20.4%, 45.6%, and 31.1% of

the eyes with a mild baseline VF defect (MD > -6 dB), respectively. Consistent with a previous study,[20] we also found that stereo photographs detected fewer eyes with glaucoma progression than GPA of the VF test in the early stages of glaucoma, but AMAF could detect progression in more eyes than GPA. These reports are comparable with previous studies using flicker chronoscopy.[21] In a previous study,[20] the agreement between stereo photographs and GPA of the VF test was moderate ($\kappa = 0.21$), as was the agreement between the AMAF ($\kappa = 0.337$) and the conventional method ($\kappa = 0.124$) in our study (Table 2). Compared to the conventional comparison method, AMAF improved agreement with the functional test and performance. It identified structural changes caused by glaucoma progression because it probably detected subtle changes of the RNFL, neuroretinal rim, and blood vessel contours (Table 3), and produced more reproducible results.

Structural progression was determined by a single grader in our study, although each image for grading was prepared by another examiner in a masked manner. To overcome any discrepancies, we compared the RNFL thickness decay of progressing eyes with that of stable eyes in each progression criteria, because the rate of structural progression should be faster in progressing eyes than in stable ones if each criterion identified the true progression caused by glaucoma (Table 4). Based upon the limited number of OCT examinations, the RNFL thickness decay

was not calculated from linear regression analysis, but from the difference between the first and the last examinations during the follow-up periods. However, the rate of average RNFL thickness decrease in our study was comparable to a previous report using the same OCT (between -2.22 μm/year and -7.60 μm/year for the Stratus OCT; Leung et al.[22]) In our series, AMAF and GPA showed greater RNFL thickness decreases in progressing eyes than stable eyes, whereas the conventional comparison method and the CIGTS did not. It is possible that the flicker technology and GPA criteria outperformed the conventional comparison and CIGTS methods in detecting glaucoma progression in patients with mild baseline VF defects. Both progressing and stable eyes, by all criteria, showed significant RNFL thickness decreases from the baseline results in our study. Liu et al.[23] reported that RNFL thicknesses decreased significantly even in eyes showing no progression, when estimated by conventional methods using both GPA of the VF test and stereo disc photography. Although the rate of RNFL decay of the study of Liu et al.[23] (between -0.71 ± 0.09 μm/year and -1.0 ± 0.20 μm/year) was less than that of our results (-1.83 ± 3.22 μm/year and -2.87 ± 3.86 μm/year in stable eyes using AMAF and GPA criteria, respectively), the use of different devices (Stratus vs. Cirrus) and decay-calculating algorithms may have influenced the results. The rate of RNFL thickness decrease as reported using Cirrus OCT may be slower than that using Stratus OCT in the same patients.[22] When GPA criterion

was combined with the conventional comparison method, RNFL thickness decreases between progressing and stable eyes were significantly different (Table 4). Based upon these results, conventional stereo photographs combined with GPA can improve the ability to assess glaucoma progression when flicker technology is not available.

Although an improvement in detection ability is an obvious advantage of the AMAF method, it does not necessarily lead to improved rates of agreement among observers, and can result in confusing results. If more information is provided than is required, it can cause confusion for inspectors who are assessing the glaucomatous progression. Confusion can involve factors involved in decreasing the agreement between readers, and decreasing agreement even for the same reader after repeated readings. Cymbor et al.[8] studied agreement among intrareader graders, and between pairs of interreader graders, when comparing the flicker and side-by-side comparison methods for the assessment of glaucomatous progression. Kappa values for intragrader agreement were 0.22 for the side-by-side comparison method and 0.46 for the flicker comparison method, and kappa values for the interreader agreement were 0.10~0.36 for the side-by-side comparison method and 0.25~0.86 for the flicker comparison method.[8] Jampel et al.[24] reported on the interobserver agreement among glaucoma experts assessing progression of optic disc changes using the side-by-side comparison method. The reported agreement

was slight to fair (κ = 0.20). Azuara-Blanco et al.[25] also investigated clinical agreement in the detection of optic disk changes in patients with glaucoma using simultaneous stereophotographs. The intraobserver agreement had a kappa value ranging from 0.55–0.78, and interobserver agreement had a kappa value ranging from 0.34–0.68. VanderBeek et al.[10] compared the agreement of digital optic disk photographs and alternating flicker for detecting progression of parapapillary atrophy. The interobserver agreement using the flicker method was better than the agreement using photographs (κ = 0.52 vs. 0.18, respectively, p = 0.02), and the intraobserver agreement was similar for both graders (photographs: κ = 0.58 vs. 0.57, p = 0.97; flicker: κ = 0.61 vs. 0.70, p = 0.37). Although there have been no previous direct comparison studies assessing the inter- and intragrader agreement for glaucomatous progression, the alternating flicker method showed better results than the classical side-by-side comparison method.

The present study compared the inter- and intragrader agreement of the AMAF and side-by-side comparison methods, using various graders (glaucoma experts and general ophthalmologists) and grading scales. Regarding interreader agreement, the AMAF method showed fair to substantial agreement for both grading scales (k = 0.530~0.645). When compared with the side-by-side comparison method, the interreader agreement of the AMAF method was better between the two glaucoma specialists, almost similar between the glaucoma

specialist and the general ophthalmologist, and worse between the two general ophthalmologists. The interreader agreement of the side-by-side comparison method between the two general ophthalmologists showed the highest kappa value and ICC coefficient among all of the interreader agreements. The same trend was found in the intrareader agreements. While both the glaucoma experts (A and B) showed a better intrareader agreement using the AMAF method for both grading scales, the general ophthalmologist (C) showed a better agreement using the classical side-by-side comparison method. When using the AMAF method to determine the structural progression of glaucoma, there was an improvement in the ability to detect a change in both the glaucoma experts and the general ophthalmologist, while there was an improvement in agreement only for the glaucoma experts., Azuara-Blanco et al.[25] reported interobserver and intraobserver agreement of glaucoma experts in the detection of optic disk changes using stereophotographs and whis was similar to the results of the present study. We postulate that the following factors were responsible for this observation. First, general ophthalmologists showed a greater tendency toward concluding that there was no structural change using both the AMAF and the side-by-side comparison methods. As shown in Table 5, general ophthalmologists were significantly more likely to reach a conclusion of “No progress” than the glaucoma experts. Due to the nature of the agreement analysis, differences in negative readings, especially

between the two general ophthalmologists, could cause differences in the agreement rate within the same group. Second, the additional information (or clues) provided by the AMAF method can cause confusion for the reader. If readers are not familiar with the assessment of subtle changes in fundus photographs and/or have little expertise of structural changes during glaucoma, it is unreasonable to expect a consistent assessment of changes in fundus photographs. For example, if there is a small change in the contour of the optic disc rim notching, it is difficult to detect this change using the side-by-side comparison method. The examiner may therefore ignore these changes in the repeated readings. Even subtle contour changes can be found using the AMAF method. There is the possibility that a general ophthalmologist might identify these changes in one reading procedure, but not in repeated readings. In addition, it is more likely that the general ophthalmologist will fail to find such subtle changes even using the AMAF methods.

Our study had some limitations. First, it used a retrospective design and a limited number of OCT examinations. The number of examinations was therefore insufficient for regression analysis to detect changes of RNFL thicknesses. Futhermorer, the optic disc photographs were interpreted by a single grader, although a masked method was used.

CONCLUSION

In conclusion, optic disc photography using a flicker chronoscopic method, AMAF, can detect more glaucoma progression than GPA of the VF in patients with a mild baseline VF defect. AMAF was better than the side-by-side comparison method in discriminating glaucomatous changes, and showed moderate agreement with functional progression using the VF test

Reference

1. Harwerth, R.S., et al., *Ganglion cell losses underlying visual field defects from experimental glaucoma.* Invest Ophthalmol Vis Sci, 1999. **40**(10): p. 2242-50.
2. Kass, M.A., 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**(6): p. 701-13; discussion 829-30.
3. Alencar, L.M., et al., *Agreement for detecting glaucoma progression with the GDx guided progression analysis, automated perimetry, and optic disc photography.* Ophthalmology, 2010. **117**(3): p. 462-70.
4. Leung, C.K., et al., *Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: patterns of retinal nerve fiber layer progression.* Ophthalmology, 2012. **119**(9): p. 1858-66.
5. Berger, J.W., et al., *Computerized stereochronoscopy and alternation flicker to detect optic nerve head contour change.* Ophthalmology, 2000. **107**(7): p. 1316-20.
6. Leske, M.C., et al., *Early Manifest Glaucoma Trial: design and baseline data.* Ophthalmology, 1999. **106**(11): p. 2144-53.
7. Heijl, A. and B. Bengtsson, *Diagnosis of early glaucoma with flicker comparisons of serial disc photographs.* Invest Ophthalmol Vis Sci, 1989. **30**(11): p. 2376-84.
8. Cymbor, M., L. Lear, and M. Mastrine, *Concordance of flicker comparison versus side-by-side comparison in glaucoma.* Optometry, 2009. **80**(8): p. 437-41.
9. Syed, Z.A., et al., *Automated alternation flicker for the detection of optic disc haemorrhages.* Acta Ophthalmol, 2012. **90**(7): p. 645-50.
10. VanderBeek, B.L., S.D. Smith, and N.M. Radcliffe, *Comparing the detection and agreement of parapapillary atrophy progression using digital optic disk photographs and alternation flicker.* Graefes Arch Clin Exp Ophthalmol, 2010. **248**(9): p. 1313-7.
11. Hodapp, E., R.K. Parrish, and D.R. Anderson, *Clinical decisions in glaucoma*. 1993,

St. Louis, Mo.: Mosby. vii, 204 p., 2 p. of plates.

12. Bay, H., T. Tuytelaars, and L.V. Gool, *SURF: Speeded Up Robust Features.* Computer Vision and Image Understanding, 2008. **110**(3): p. 346-359.
13. Musch, D.C., et al., *Visual field progression in the Collaborative Initial Glaucoma Treatment Study the impact of treatment and other baseline factors.* Ophthalmology, 2009. **116**(2): p. 200-7.
14. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.
15. Radcliffe, N.M., et al., *Comparison of stereo disc photographs and alternation flicker using a novel matching technology for detecting glaucoma progression.* Ophthalmic Surg Lasers Imaging, 2010. **41**(6): p. 629-34.
16. Bengtsson, B. and C.E. Krakau, *Flicker comparison of fundus photographs. A technical note.* Acta Ophthalmol (Copenh), 1979. **57**(3): p. 503-6.
17. Syed, Z.A., et al., *Detection of progressive glaucomatous optic neuropathy using automated alternation flicker with stereophotography.* Arch Ophthalmol, 2011. **129**(4): p. 521-2.
18. Malik, R., W.H. Swanson, and D.F. Garway-Heath, *'Structure-function relationship' in glaucoma: past thinking and current concepts.* Clin Experiment Ophthalmol, 2012. **40**(4): p. 369-80.
19. Harwerth, R.S., et al., *Linking structure and function in glaucoma.* Prog Retin Eye Res, 2010. **29**(4): p. 249-71.
20. Banegas, S.A., et al., *Agreement among spectral-domain optical coherence tomography, standard automated perimetry, and stereophotography in the detection of glaucoma progression.* Invest Ophthalmol Vis Sci, 2015. **56**(2): p. 1253-60.
21. Chee, R.I., et al., *Agreement of flicker chronoscopy for structural glaucomatous progression detection and factors associated with progression.* Am J Ophthalmol, 2013. **155**(6): p. 983-990 e1.
22. Leung, C.K., et al., *Evaluation of retinal nerve fiber layer progression in glaucoma: a comparison between spectral-domain and time-domain optical coherence tomography.* Ophthalmology, 2011. **118**(8): p. 1558-62.

23. Liu, T., et al., *Rates of Retinal Nerve Fiber Layer Loss in Contralateral Eyes of Glaucoma Patients with Unilateral Progression by Conventional Methods.* Ophthalmology, 2015. **122**(11): p. 2243-51.
24. Jampel, H.D., et al., *Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients.* Am J Ophthalmol, 2009. **147**(1): p. 39-44 e1.
25. Azuara-Blanco, A., et al., *Clinical agreement among glaucoma experts in the detection of glaucomatous changes of the optic disk using simultaneous stereoscopic photographs.* Am J Ophthalmol, 2003. **136**(5): p. 949-50.

국문요약

목적 : 녹내장 환자에서 새롭게 개발된 AMAF (automatic matched matched flicker)의 진행 탐지 알고리즘을 평가하고자 한다.

방법 : 시야 검사에서 평균 편위값(Mean deviation) 이 −6 dB보다 큰 103명의 개방각 녹내장 환자를 대상으로 후향적 연구를 진행하였다. 기능적 진행은 시야검사에 의해 판별되는 두가지의 진행기준을 근거로 판단되었으며, 구조적인 진행은 시신경 유두 및 망막 신경 섬유층 (RNFL) 촬영을 사용하여 기존의 병렬비교법과 AMAF를 통한 판단법을 이용하였다. AMAF와 기존의 병렬비교법의 녹내장 진행 판별능력의 비교는 기능적 및 구조적 진행 판독 결과간의 일치도에 의해 평가되었다. 빛간섭단층촬영 (OCT, optical coherence tomography) 을 통해 측정된 RNFL 두께 변화가 상기의 각각의 판독법을 통해 진행 또는 비진행으로 판독된 안구에서 어떤 차이를 보이는지도 비교하였다. 상기 연구에 포함되지 않았던 45쌍의 정상안압녹내장 환자의 안저사진을 이용해 일치도 연구를 진행하였다. 병렬 비교 이미지 그리고 AMAF로 만들어진 교대깜빡임 이미지를 두명의 녹내장 전문의와 두명의 일반 안과의사가 판독하였다. 검사자내 그리고 검사자간 일치도를 카파상수와 Bland and Altman plot을 이용해 분석하였다.

결과 : 103안중 47 (45.6%), 21 (20.4%), 그리고 32 (31.1%)안이 각각 AMAF, 기존의 병렬비교법 그리고 시야검사를 이용한 guided progression analysis (GPA) 를 통해 녹내장이 진행된 것으로 판독되었다. AMAF 는 기존에 병렬비교법에 비해 시야검사의 GPA 와 일치도

가 더 우수한 것으로 나타났다. (κ = 0.337; P < 0.001 and κ = 0.124; P = 0.191, respectively). OCT를 통해 측정된 시신경섬유층 두께의 변화는 AMAF를 통해서 진행된 녹내장으로 판명된 환자군에서 진행이 없는 것으로 판명된 환자군에 비해 유의하게 컸던 반면 (−3.49 ± 2.86 μm/year vs. −1.83 ± 3.22 μm/year; P = 0.007) 병렬비교법에서는 그렇지 못했다. (−3.24 ± 2.42 μm/year vs. −2.42 ± 3.33 μm/year; P = 0.290).

검사자간 일치도 비교 연구에서, 녹내장 전문의는 AMAF를 통해서 (average, 40%) 기존의 병렬비교법에 비해 (average, 23.3%) 더 많은 녹내장 진행을 판독해 내었다. 일반 안과의사 역시 AMAF를 이용해 (average, 28.9%) 기존의 병렬비교법보다 (average, 18.9%) 더 많은 녹내장을 판독해 내었다. AMAF 는 평균이상에서 상당한 정도의 (fair to substantial) 검사자간 일치도를 (k = 0.530~0.645) 평균에서 완벽한 검사자내 일치도 (k = 0.503~0.955) 를 보였다. AMAF를 이용한 녹내장 진행의 판독에 있어서 검사자간 그리고 검사자내 일치도는 녹내장 전문의가 일반 안과의사에 비해 우수하였다.

결론 : AMAF는 녹내장의 구조적 진행을 발견해 내는데 있어서 기존의 병렬비교법에 비해 더 우수한 것으로 확인되었으며 시야검사를 통한 기능적 진행검사와도 중등도의 일치도를 보였다. 또한 녹내장 전문의가 이용하였을 때 우수한 검사자간 검사자내 일치도를 보여서 녹내장 환자의 진행 관찰에 도움이 될 수 있을 것으로 사료된다.

Figure 1. A. The appearance of the application. B. When examiner check “User mask” to ‘Yes’ on flicker applications, examiner can mark area that examiners want to make it a reference area for image fusion. C. Result of matching point after examiner marks area that examiners want to make it a reference area.

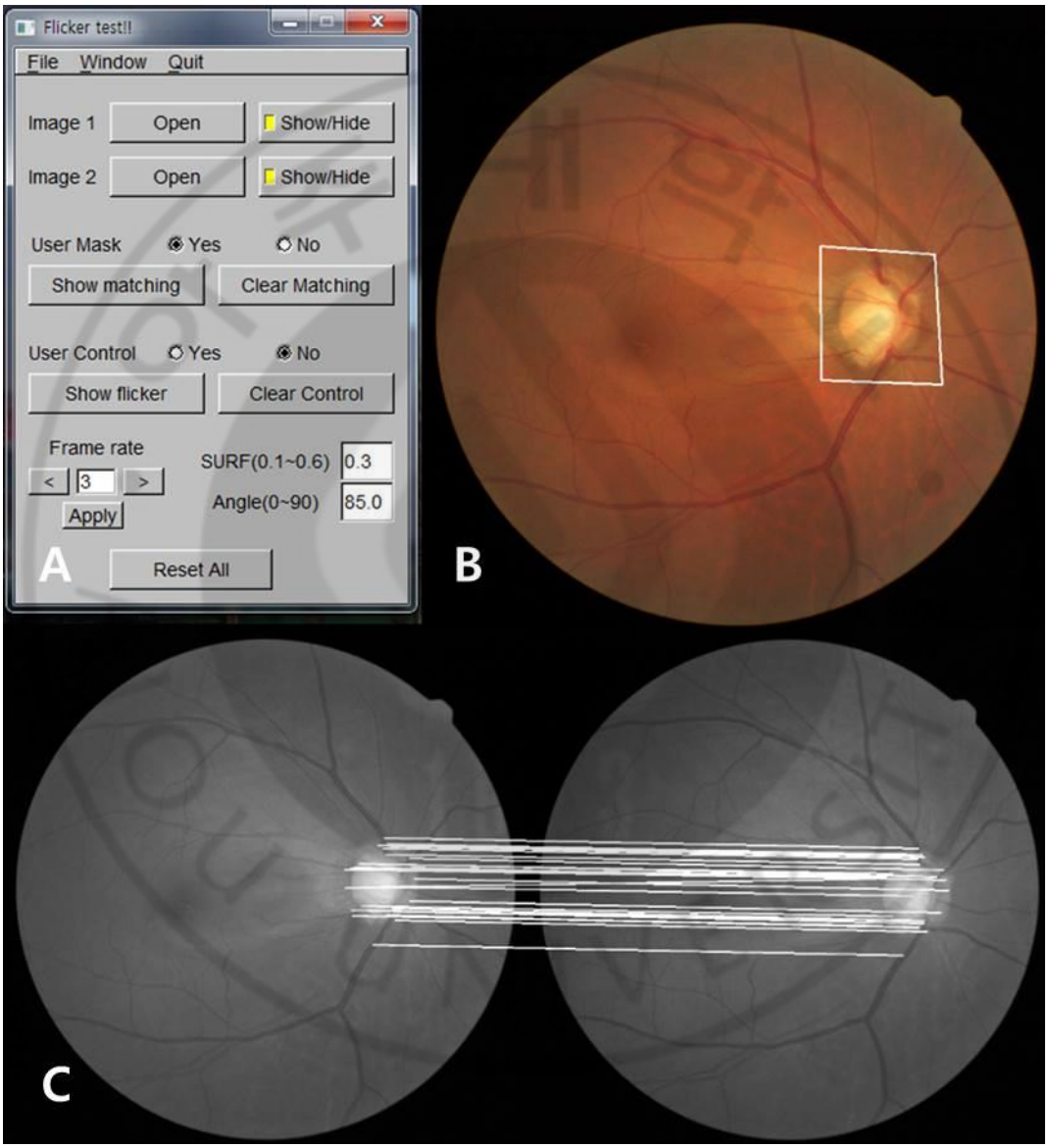

Figure 2. A. The example of screen view that underwent progression analysis using automated matched alternating flicker. B. The example of screen view that underwent progression analysis using classical side-by-side comparison.

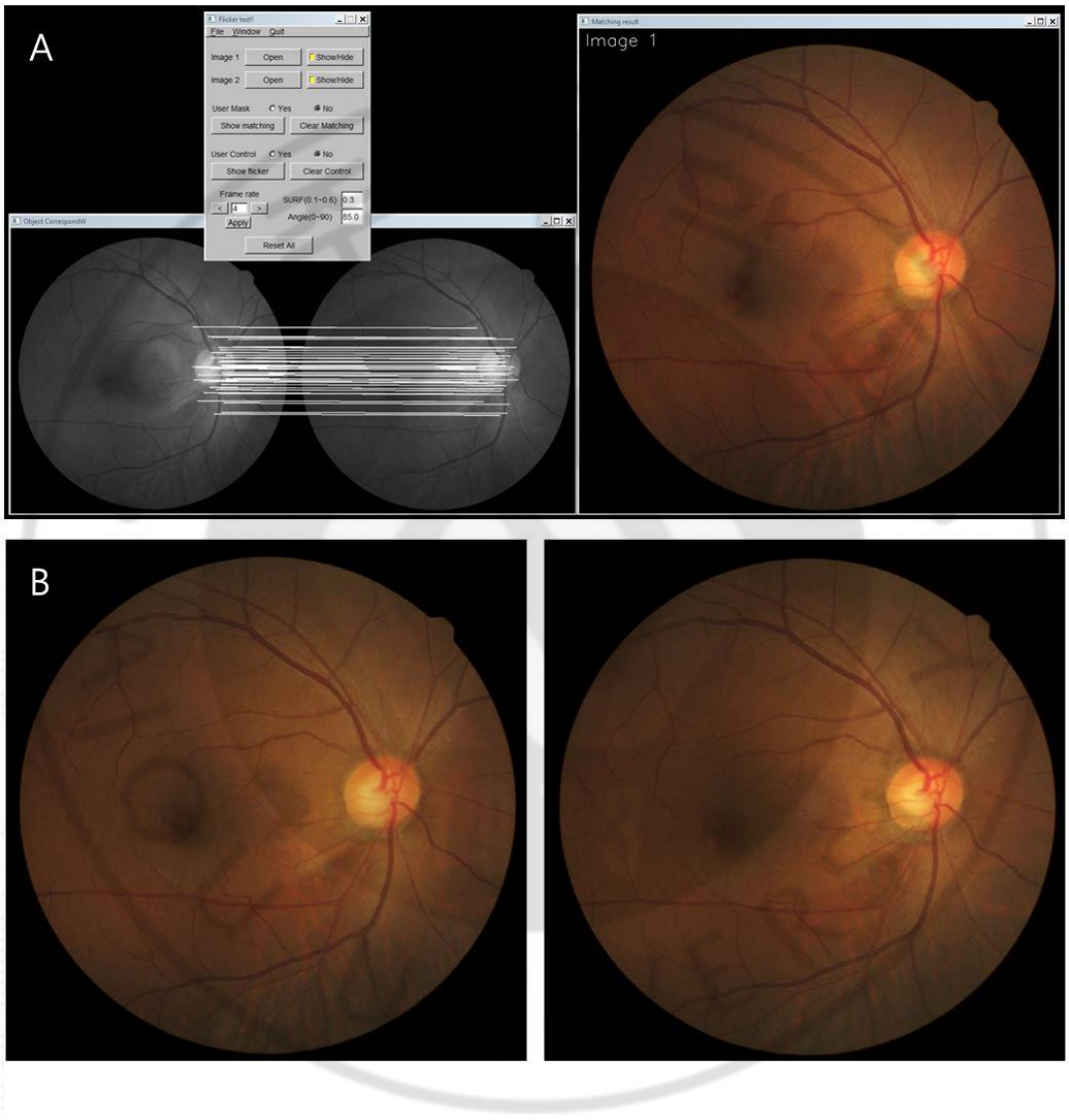

Table 1. Descriptive statistics of enrolled subjects (means ± SD)

	Baseline examination	Last examination	p value
Duration		28.7 ± 9.0 months	
MD, dB	-0.93 ± 2.14	-1.62 ± 2.93	< 0.001
PSD, dB	3.97 ± 2.79	4.52 ± 3.41	< 0.001
CIGTS score	2.06 ± 2.43	2.78 ± 2.93	< 0.001
RNFL thickness measured by OCT (μm)			
Average	87.79 ± 12.78	81.86 ± 13.79	< 0.001
Temporal quadrant	69.84 ± 13.81	65.85 ± 13.15	< 0.001
Superior quadrant	107.38 ± 20.35	100.64 ± 22.19	< 0.001
Nasal quadrant	71.83 ± 15.16	68.17 ± 15.00	< 0.001
Inferior quadrant	101.02 ± 25.38	92.82 ± 25.15	< 0.001

Statistical significance was evaluated by paired T-test

MD, mean deviation; PSD, pattern standard deviation; CIGTS, collaborative initial glaucoma treatment study; RNFL, retinal nerve fiber layer; OCT, optical coherence tomography

Table 2. Agreement between image analysis methods and VF criteria to detect glaucoma progression

	GPA progression criteria			CIGTS progression Criteria		
	Stable	Progress	κ (p-value)	Stable	Progress	κ (p-value)
AMAF						
Stable	47	9	0.337	51	5	0.241
Progress	24	23	(<0.001)	32	15	(0.003)
Conventional methods						
Stable	59	23	0.124	69	13	0.178
Progress	12	9	(0.191)	14	7	(0.071)

GPA, glaucoma progression analysis; CIGTS, collaborative initial glaucoma treatment study; κ, Kappa coefficient; AMAF, automated matched alternating flicker

Table 3. Agreement between matched flicker and conventional comparison method to detect changes of fundus photography.

			Conventional comparison method			
			Stable	Progress	Kappa	p-value
RNFL defect widening	AMAF	Stable	75	0	0.785	<0.001
		Progress	8	20		
RNFL defect depth deepening	AMAF	Stable	74	0	0.354	<0.001
		Progress	21	8		
Rim contour change	AMAF	Stable	76	1	0.241	0.001
		Progress	21	5		
Vessel contour change	AMAF	Stable	88	0	0.109	0.015
		Progress	14	1		

RNFL, retinal nerve fiber layer; AMAF, automated matched alternating flicker

Table 4. Comparison of retinal nerve fiber layer decay between progressing and stable eyes according to the progression criteria.

Progression criteria	RNFL decrement (μm/year)		*p*-value*
	Progressing eyes	Stable eyes	
Conventional methods	-3.24 ± 2.42	-2.42 ± 3.33	0.290
AMAF	-3.49 ± 2.86	-1.83 ± 3.22	0.007
GPA	-3.81 ± 3.20	-2.87 ± 3.86	0.008
CIGTS	-3.63 ± 2.24	-2.34 ± 3.31	0.101
Conventional methods combined with GPA	-3.56 ± 2.98	-1.86 ± 3.12	0.006
AMAF combined with GPA	-3.52 ± 3.02	-1.47 ± 2.99	0.001

* independent student t-test between progressing and stable eyes

RNFL, retinal nerve fiber layer; AMAF, automated matched alternating flicker; GPA, glaucoma progression analysis; CIGTS, collaborative initial glaucoma treatment study

Table 5. Summary of Grading of glaucomatous progression by Glaucoma experts (A, B) and general ophthalmologist(C, D) using with automated matched alternating flicker(AMAF) and side-by-side parallel comparison.

		AMAF				side-by-side comparison			
		A	B	C	D	A	B	C	D
5-point scale	1	13 (28.9%)	18 (40.0%)	17 (37.8%)	27 (60.0%)	19 (42.2%)	29 (64.4%)	29 (64.4%)	33 (73.3%)
	2	4 (8.9%)	0 (0 %)	1 (2.2%)	0 (0.0%)	12 (26.7%)	0 (0.0%)	2 (4.4%)	1 (2.2%)
	3	6 (13.3%)	3 (6.7%)	7 (15.6%)	2 (4.4%)	3 (6.7%)	2 (4.4%)	4 (8.9%)	1 (2.2%)
	4	7 (15.6%)	3 (6.7%)	7 (15.6%)	3 (6.7%)	4 (8.9%)	0 (0.0%)	2 (4.4%)	1 (2.2%)
	5	15 (33.3%)	21 (46.7%)	13 (28.9%)	13 (28.9%)	7 (15.6%)	14 (31.1%)	8 (17.8%)	9 (20.0%)
yes or no scale	No	23 (51.1%)	21 (46.7%)	25 (55.6%)	29 (64.4%)	34 (75.6%)	30 (66.7%)	35 (77.8%)	35 (77.8%)
	Yes	22 (48.9%)	24 (53.3%)	20 (44.4%)	16 (35.6%)	11 (24.4%)	15 (33.3%)	10 (22.2%)	10 (22.2%)

Table 6. Intra-reader and inter-reader agreement for evaluation of glaucomatous fundus changes with automated matched alternating flicker (AMAF) and side-by-side parallel comparison. [Kappa value (ICC coefficient)]

Inter-reader agreement [Kappa value(ICC coefficient)]			
		AMAF	side-by-side comparison
5-point scale	A – B (Between two Glaucoma experts)	0.584 (0.782)	0.490 (0.724)
	A – C (Expert and general ophthalmologist)	0.556 (0.764)	0.580 (0.865)
	C – D (Between two general ophthalmologist)	0.530 (0.760)	0.821 (0.952)
	Among 4 readers	N/A (0.871)	N/A (0.913)
yes or no scale	A – B (Between two Glaucoma experts)	0.645 (0.786)	0.464 (0.642)
	A – C (Expert and general ophthalmologist)	0.554 (0.715)	0.690 (0.817)
	C – D (Between two general ophthalmologist)	0.541 (0.709)	0.871 (0.931)
	Among 4 readers	N/A (0.856)	N/A (0.894)
Intra-reader agreement			
	AMAF	side-by-side comparison	
5-point scale	A	0.951 (0.992)	0.762 (0.917)
	B	0.910 (0.966)	0.823 (0.965)
	C	0.548 (0.769)	0.833 (0.884)
yes or no scale	A	0.911 (0.955)	0.759 (0.863)
	B	0.955 (0.978)	0.800 (0.890)
	C	0.503 (0.669)	0.938 (0.969)

Table 7. Intra-instrument agreement for evaluation of glaucomatous fundus changes with automated matched alternating flicker (AMAF) and side-by-side parallel comparion.[Kappa value(ICC coefficient)]

	A	B	C	D
5-point scale	0.448(0.683)	0.506(0.663)	0.464(0.715)	0.699(0.843)
yes or no scale	0.416(0.592)	0.435(0.611)	0.526(0.694)	0.682(0.814)