Biomedical Optics EXPRESS

High-speed, image-based eye tracking with a scanning laser ophthalmoscope

Christy K. Sheehy,[1,*] Qiang Yang,[2] David W. Arathorn,[2] Pavan Tiruveedhula,[1] Johannes F. de Boer,[3] and Austin Roorda[1]

[1]*School of Optometry, University of California, Berkeley; Berkeley, CA 94720, USA*
[2]*Montana State University, Bozeman, MT 59717, USA*
[3]*Rotterdam Ophthalmic Institute, Rotterdam, The Netherlands*
**Christy.Sheehy@berkeley.edu*

Abstract: We demonstrate a high-speed, image-based tracking scanning laser ophthalmoscope (TSLO) that can provide high fidelity structural images, real-time eye tracking and targeted stimulus delivery. The system was designed for diffraction-limited performance over an 8° field of view (FOV) and operates with a flexible field of view of 1°–5.5°. Stabilized videos of the retina were generated showing an amplitude of motion after stabilization of 0.2 arcmin or less across all frequencies. In addition, the imaging laser can be modulated to place a stimulus on a targeted retinal location. We show a stimulus placement accuracy with a standard deviation less than 1 arcmin. With a smaller field size of 2°, individual cone photoreceptors were clearly visible at eccentricities outside of the fovea.

© 2012 Optical Society of America

OCIS codes (170.4460) Ophthalmic optics and devices; (330.2210) Vision - eye movements; (330.4460) Ophthalmic optics and devices.

References and links

1. S. R. Sadda, P. A. Keane, Y. Ouyang, J. F. Updike, and A. C. Walsh, "Impact of scanning density on measurements from spectral domain optical coherence tomography," Invest. Ophthalmol. Vis. Sci. **51**(2), 1071–1078 (2010).
2. E. Garcia-Martin, I. Pinilla, E. Sancho, C. Almarcegui, I. Dolz, D. Rodriguez-Mena, I. Fuertes, and N. Cuenca, "Optical coherence tomography in retinitis pigmentosa: reproducibility and capacity to detect macular and retinal nerve fiber layer thickness alterations," Retina **32**(8), 1581–1591 (2012).
3. L. C. Sincich, Y. Zhang, P. Tiruveedhula, J. C. Horton, and A. Roorda, "Resolving single cone inputs to visual receptive fields," Nat. Neurosci. **12**(8), 967–969 (2009).
4. C. R. Vogel, D. W. Arathorn, A. Roorda, and A. Parker, "Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy," Opt. Express **14**(2), 487–497 (2006).
5. D. W. Arathorn, Q. Yang, C. R. Vogel, Y. Zhang, P. Tiruveedhula, and A. Roorda, "Retinally stabilized cone-targeted stimulus delivery," Opt. Express **15**(21), 13731–13744 (2007).
6. Q. Yang, D. W. Arathorn, P. Tiruveedhula, C. R. Vogel, and A. Roorda, "Design of an integrated hardware interface for AOSLO image capture and cone-targeted stimulus delivery," Opt. Express **18**(17), 17841–17858 (2010).
7. A. Gómez-Vieyra, A. Dubra, D. Malacara-Hernández, and D. R. Williams, "First-order design of off-axis reflective ophthalmic adaptive optics systems using afocal telescopes," Opt. Express **17**(21), 18906–18919 (2009).
8. S. Poonja, S. Patel, L. Henry, and A. Roorda, "Dynamic visual stimulus presentation in an adaptive optics scanning laser ophthalmoscope," J. Refract. Surg. **21**(5), S575–S580 (2005).
9. J. B. Mulligan, "Recovery of motion parameters from distortions in scanned images," in *Proceedings of the NASA Image Registration Workshop (IRW97)* (NASA Goddard Space Flight Center, MD, 1997), no. 19980236600
10. American National Standard for the Safe Use of Lasers, ANSI Z136.1–2007 (Laser Institute of America, Orlando, 2007)
11. S. B. Stevenson, A. Roorda, and G. Kumar, "Eye tracking with the adaptive optics scanning laser ophthalmoscope" in *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications,*S.N. Spencer, ed. (Association for Computed Machinery, New York, 2010), pp. 195–198.
12. R. Engbert and R. Kliegl, *In The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*, J. Hyona, R. Radach, and H. Deubel, eds. (Elsevier, Oxford, 2003).103–117.
13. S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," Nat. Rev. Neurosci. **5**(3), 229–240 (2004).

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

14. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," Opt. Lett. **33**(2), 156–158 (2008).
15. L. A. Riggs, J. C. Armington, and F. Ratliff, "Motions of the retinal image during fixation," J. Opt. Soc. Am. **44**(4), 315–321 (1954).
16. L. A. Riggs and A. M. Schick, "Accuracy of retinal image stabilization achieved with a plane mirror on a tightly fitting contact lens," Vision Res. **8**(2), 159–169 (1968).
17. T. N. Cornsweet and H. D. Crane, "Accurate two-dimensional eye tracker using first and fourth Purkinje images," J. Opt. Soc. Am. **63**(8), 921–928 (1973).
18. H. D. Crane and C. M. Steele, "Generation-V dual-Purkinje-image eyetracker," Appl. Opt. **24**(4), 527–537 (1985).
19. F. Santini, G. Redner, R. Iovin, and M. Rucci, "EyeRIS: a general-purpose system for eye-movement-contingent display control," Behav. Res. Methods **39**(3), 350–364 (2007).
20. M. Rucci, R. Iovin, M. Poletti, and F. Santini, "Miniature eye movements enhance fine spatial detail," Nature **447**(7146), 852–854 (2007).
21. E. Midena, "Liquid crystal display microperimetry," in *Perimetry and the Fundus: an Introduction to Microperimetry*, E. Midena, ed. (Slack, Thorofare, NJ, 2007), pp. 15–26.
22. D. X. Hammer, R. D. Ferguson, C. E. Bigelow, N. V. Iftimia, T. E. Ustun, and S. A. Burns, "Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging," Opt. Express **14**(8), 3354–3367 (2006).
23. S. Martinez-Conde, S. L. Macknik, X. G. Troncoso, and D. H. Hubel, "Microsaccades: a neurophysiological analysis," Trends Neurosci. **32**(9), 463–475 (2009).

1. Introduction

The human eye is constantly in motion. Even when fixating on a target, our eyes move; drifting and making microsaccades which move a stimulus projected onto the retina over dozens to hundreds of photoreceptors. With the eye as an ever moving target, our ability to record high-fidelity images of the retina is limited. Moreover, targeted light delivery to the retina remains uncontrolled with constant eye motion.

Recent advances in imaging technology have highlighted the importance of improved eye tracking to render true and accurate images. In the clinical domain, active eye tracking has proven to be effective in commercial systems [1,2]. At a more basic level, the benefits of accurate eye tracking and stimulus delivery prove to be useful for delivering stimuli to targeted retinal locations as small as a single cone [3].

An image-based method for eye tracking and targeted stimulus delivery has been implemented into an adaptive optics scanning laser ophthalmoscope (AOSLO) system in our lab and has been reported in a series of publications [4–6]. In this paper, we show that the same image-based tracking techniques can be implemented in a more traditional, larger field of view, confocal SLO. This system is the most accurate, fast and functional tracking system to be used in a standard ophthalmic instrument and demonstrates that rich texture in the image, not necessarily the presence of cones, is sufficient for this tracking method. The use of a conventional approach offers a more robust, compact and cost effective system that is readily deployable in a variety of settings. We will show that in a well-designed SLO system, the wider field of view (FOV) is able to capture retinal video rich with structure, allowing accurate image-based tracking during normal fixational eye movements.

2. Methods

2.1. System hardware

The tracking scanning laser ophthalmoscope (TSLO) was developed in the following manner. The optical design and system optimization was completed using optical design software (Radiant ZEMAX LLC, Bellevue, WA). System specifications were as follows:

- Diffraction-limited optical design over an 8° FOV (excluding the eye)
- Adjustable pupil size between 2 and 4 mm (no need for subject dilation)
- Small focal length mirrors for a compact design
- Flexible eye relief

The main portion of the system contains three telescope assemblies that relay the pupil to the fast and slow scan mirrors and then to the light detection and delivery arm (Fig. 1). The

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

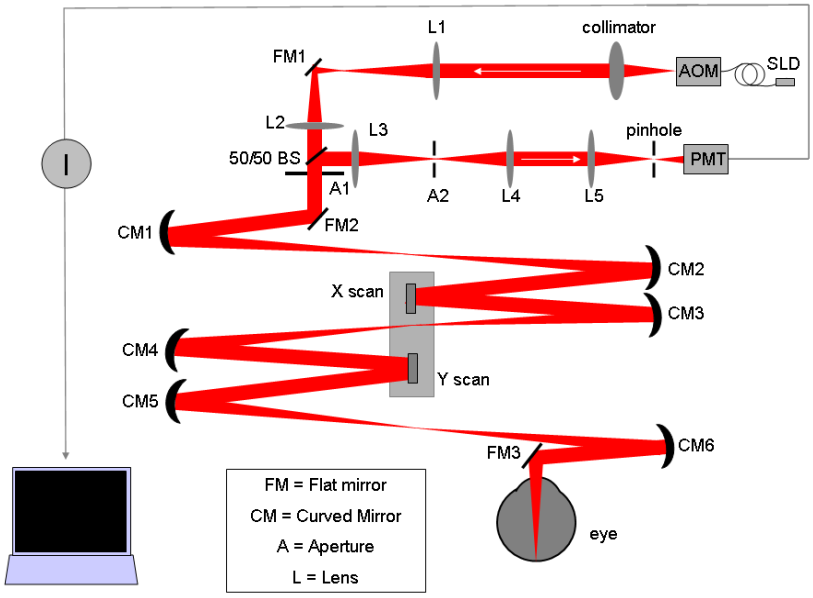

Fig. 1. A 2-D optical design schematic of the TSLO. Light exiting the super luminescent diode (SLD) is coupled into the acousto-optic modulator (AOM) before entering the system. The light is collimated and sent through a basic 4f series of lenses onto an adjustable aperture (A1). Light travels through three mirror based telescope assemblies (f = 250 mm) to the human eye. Light is then reflected off the retina and sent back through the system into the light detection arm. Another series of lenses in a 4f configuration relays the light to be collected by a photomultiplier tube (PMT). A 50 µm pinhole (1.95 Airy disc diameters for a 4mm pupil) is placed at the retinal conjugate plane prior to the PMT for confocality. The intensity (I) of the signal is sent to the PC for readout. This is a schematic layout, the actual components are not aligned in a single plane (see Fig. 2).

telescopes are arranged in such a way so as to minimize astigmatism in both the pupil and retinal planes [7]. Each concave mirror used has a focal length of f = 250 mm. A pinhole was placed at the retinal conjugate prior to the photomultiplier tube (PMT) in order to make the system confocal. A galvo scanner and resonant scanner were placed into the system at pupil conjugate positions to scan the beam across the subject's retina both vertically and horizontally, respectively. The horizontal, or fast, scanner operates at ~16 kHz, while the vertical scanner operates at a rate of 1/512 of the fast scan to record frames at ~30 frames per second. An image is created, pixel by pixel, with each frame consisting of 512 x 512 pixels. Since each video frame is acquired over time, there are a unique set of distortions created by the subject's eye motion. It is these distortions that are used to extract the motion of the eye in real time.

The opto-mechanical design for the system components was modeled in Solidworks (Concord, MA). Appropriate heights for mounting were determined based on the 3-D drawing. The use of Solidworks to lay out an optical system proved extremely helpful, as one can export the beam path directly from ZEMAX into the program and determine the necessary opto-mechanical components (Fig. 2).

Diffraction-limited performance was achieved for nine points over an 8 degree FOV. The geometric spots at the nine points across the field are well within the Airy disc (represented by the circles), with most spots less than 5 µm (Fig. 3).

An 840 nm super luminescent diode (SLD) (Superlum, Moscow, Russia), with a 50 nm bandwidth, was used for imaging. The light source is connected to an acousto-optic modulator (AOM) whose output intensity is continuously controlled with a voltage output from a 14-bit digital to analog converter (DAC). As the beam raster-scans the retina, the DAC drives the AOM to modulate the SLD so that it is only on during the central 80% of the forward sweep

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

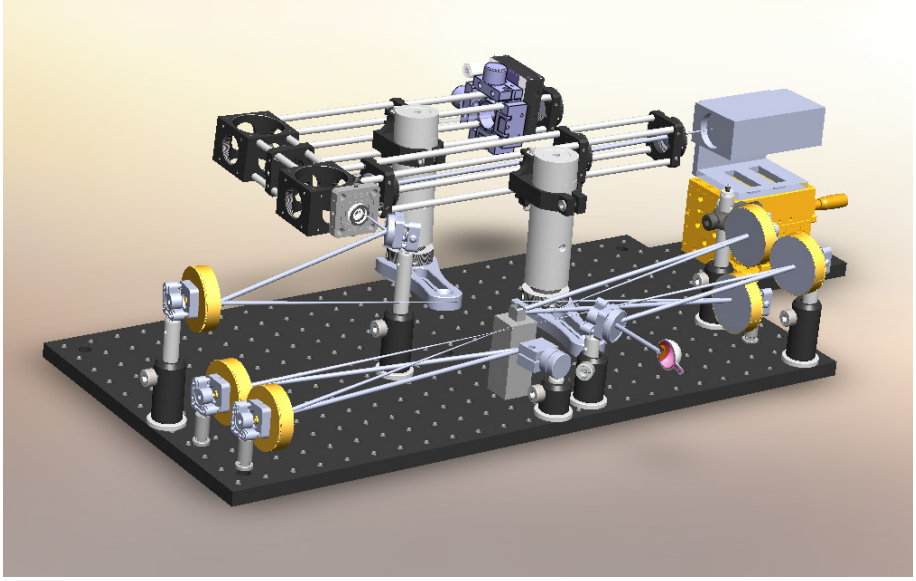

Fig. 2. Opto-mechanical design (to scale) of the TSLO on a 60 cm x 30 cm breadboard (length x width). The telescope assemblies are built in such a way so as to limit system astigmatism through the varying of beam heights and angles. Note the changing of beam heights as the light propagates through the system to the eye. Light source, acousto-optic modulator (AOM) and chin rest not shown.

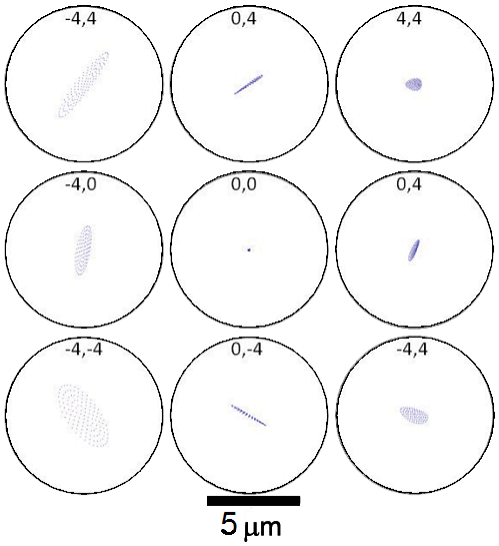

Fig. 3. Geometric spot diagrams contained within the calculated Airy discs at nine points spanning the 8° FOV (λ = 840 nm with a pupil size of 4 mm). The coordinates represent the field location in degrees.

of the mirror scanning cycle, thereby limiting the exposure to only those times when the light is being detected. The AOM is also used to modulate SLD power to place any gray-scale image point by point onto the retina [8]. A stimulus presented in this way appears in negative contrast (i.e. laser is switched off to write the stimulus) within the dim red field created by the scanning light source. Since the AOM is synchronized with the scanning of the beam, modulation timing can be manipulated to place a stimulus at any location within the raster scan [4]. In this manner, the imaging and stimulus delivery are done with the same light

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

source. Other variants of this system using secondary sources have been reported [3], but are not used here.

2.2. Software and hardware for eye tracking

The details of the eye motion recovery used in this system have been previously reported [5,6,9]. Briefly, we will describe the methods herein. In order to extract retinal motion from the scanned images in real time, each frame of an SLO movie is broken up into a set number of strips that are parallel to the fast scanner. The number of strips is flexible and can be changed according to the user's experimental requirements. For each movie, a reference frame is selected, usually the first frame to occur in the series unless otherwise reselected. Each strip within a given frame is then cross-correlated with the reference frame. The (x,y) displacements of the new frame with respect to the reference frame are a measure of the relative motion of the eye at that specific point in time. Every subsequent frame can then be redrawn to align it with the reference frame (Fig. 4). This occurs in real-time so that the operator can see both the subject's actual retinal motion and the stabilized version of the retina side by side on the software interface. Using the real-time eye trace generated from the (x,y) displacements of each frame as described above, the timing of the stimulus delivery can be controlled to guide its placement to any targeted location on the retina.

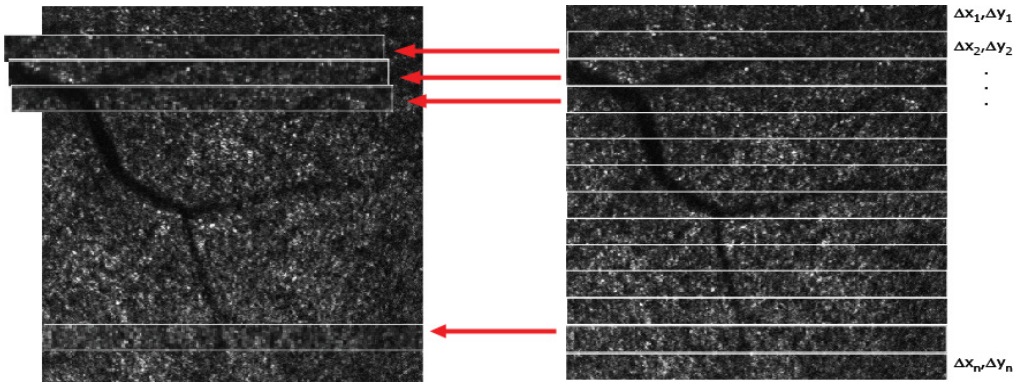

Fig. 4. Recovery of eye motion. Each frame is broken into strips that are cross correlated with the reference frame (left) in order to determine the (x,y) displacements. In the current implementation, these cross-correlations are made to the nearest pixel.

For the data presented here, cross-correlations are performed and (x,y) displacements are reported 32 times per frame, for a reporting rate of 960 Hz. The correlations are computed from 32 overlapping strips per frame, each of which is 32 pixels high.

Eye position estimates are made after-the-fact (i.e. after the strip has been recorded). This computation is done very quickly, but the delivery of a stimulus to a targeted retinal location requires a prediction. Figure 5 shows the steps involved. For a 16 kHz line scan rate, the latency is 2.5 ± 0.5 msec, depending on where in the 16 pixel strip the stimulus actually starts.

To ensure that the targeted stimulus delivery is accurate, we impose a threshold on the cross-correlation peak. In the current system, whenever the normalized cross-correlation peak is less than 0.3, a decision is made to not deliver the stimulus. Reductions in the cross-correlation peak will occur whenever (i) the amount of overlap between the current strip and the reference is reduced (mainly due to horizontal eye movements), (ii) the features within the strip are distorted because of eye motion (iii) the quality of the image is reduced due to tear film break-up, accommodation, or pupil constriction (iv) the image is lost due to blinks, (v) the retinal image changes because of lateral and axial pupil displacements relative to the scanning beam and (vi) there are intrinsic changes in the reflectivity of the retina.

The entire process is computationally demanding and requires a custom solution. In this version, the TSLO data is recorded with a custom-programmed field programmable gate array (FPGA) board [6]. The FPGA board (Xilinx, San Jose, CA) allows for the immediate access

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

to the strips of image data as they are acquired. The cross correlation with the reference image employs a standard FFT-based algorithm and takes place on a graphics board (Nvidia, Santa Clara, CA) on the host PC.

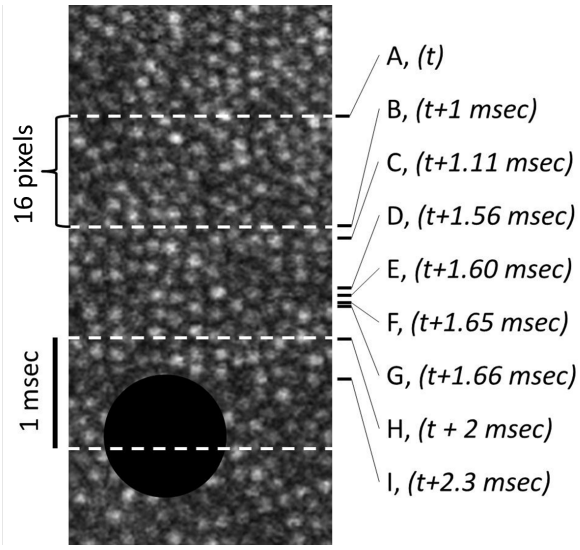

Fig. 5. Timelines for targeted stimulus delivery. The figure shows a cropped region of an SLO frame split into 16 pixel segments as indicated by the dashed white lines. With a 16 kHz line rate, each 16 pixel segment takes 1 msec to complete as indicated on the right side of the figure. The right indicators show time points for the critical tasks involved in delivering a stimulus, in this case a black circle, to a targeted location. A, Center location of the 32-pixel image strip where position estimate is to be made. B, The final line in the current 32-pixel image strip is scanned and the complete strip is recorded into the FPGA memory. C, The image strip is read into the PC/GPU. D, The eye position at time *t* relative to the reference frame is estimated by the GPU. E, The time from *t* to the first line of the buffer that contains the estimated position of the stimulus is computed. F, Make decision: Is there enough time to wait for the next strip before arming the AOM buffer with the stimulus? If YES, then wait for next strip and repeat process. If NO, then proceed to step G. G, Load AOM buffer so it is armed to play out the stimulus while the beam scans over the target. H, Start playing out the AOM buffer at this strip. I, Start delivering the stimulus (note: when the stimulus is greater than 64 pixels in size, steps B to I are repeated for every strip that the stimulus occupies).

2.3. Testing on model and human eyes

Detailed tracking performance was quantified on a custom-built model eye, which used a galvo scanner mirror placed between the optics and the retina. This allowed for controlled amounts of retinal motion with fixed frequencies and amplitudes.

Human eye data are reported here for two subjects who are also co-authors on this paper. The experiment was approved by the University of California, Berkeley, Committee for the Protection of Human Subjects and all protocols adhered to the tenets of the Declaration of Helsinki. A chin rest with temple pads was used in order to minimize head motion for all human eye experiments.

All measurements were recorded with a 4° FOV (512 x 512 pixels) providing a sampling resolution of 0.47 arcminutes per pixel. The power of the 840 nm light source never exceeded 500 μW at the pupil plane, which was computed to be within the ANSI safety limits [10].

3. Results

3.1. Frequency analysis

To quantify motion reduction as a function of frequency, sinusoidal retinal motion was input into the model eye. Both a raw video and stabilized video were recorded at each input

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

frequency. An offline stabilization program was used to compute retinal motion in the raw and stabilized video. The offline software generated eye motion estimates at a frequency of 1920 Hz (64 strips per frame). The amplitude spectra were then computed for both eye motion traces and the amplitudes of each spectrum at the input frequency were compared before and after stabilization. In this manner, any residual motion in the stabilized video that was not corrected by the real-time system could be measured. Eye motion measurements are recorded at nearly 100% accuracy for frequencies up to 32 Hz (Fig. 6). The estimated bandwidth for 50% correction based on a double exponential fit to the data was just over 400Hz.

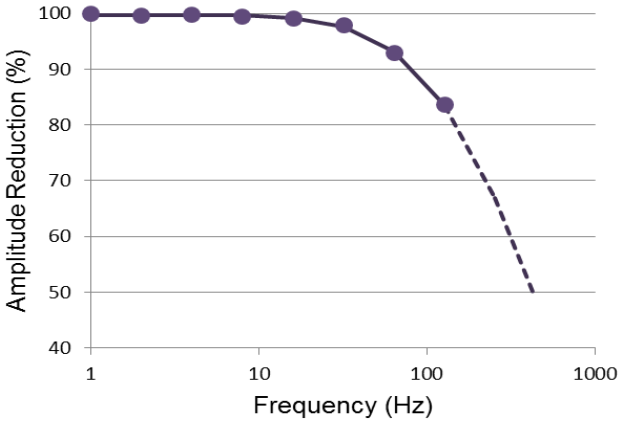

Fig. 6. Motion amplitude reduction as a function of input frequency for the model eye. For input frequencies just over 100 Hz, the reduction in amplitude is still over 80%.

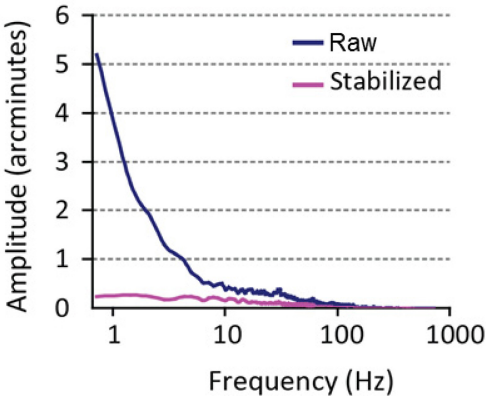

Fig. 7. Amplitude of eye motion vs. input frequency. The blue curve represents the actual motion as a function of frequency during normal fixation recorded in the raw video. The pink line represents the residual motion of features in the retinal video after stabilization. The resultant amplitude with stabilization is 0.2 arcminutes or less across all frequencies.

Next, raw and stabilized videos were taken of the human retina in real time. Ten videos of ten seconds each were recorded of one of the subjects. The same offline stabilization program with increased sampling was used for the human eye videos. The percentage of erroneous eye motion estimates was 0.83%. The standard deviation of the residual motion of features in the stabilized videos was 0.19 minutes of arc (0.41 pixels). The amplitude spectra of motion as a function of frequency shows how the actual motion in the human eye is corrected in this system. Figure 7 shows two important results. First, eye motion of a normal eye fixating on a target is dominated by low frequencies, with the amplitude dropping proportionally with the inverse of the frequency [11]. At frequencies greater than 10 Hz, the amplitude of normal fixational eye motion is less than 0.5 arcminutes. Second, the TSLO suppresses the eye

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

motion during normal fixation up to 100 Hz. The suppression of eye motion in the stabilized video means that eye motion is being reliably measured over these frequencies in real-time.

3.2. Threshold velocity of eye motion

The system's maximum tolerable velocity was computed in order to determine how fast an eye movement can be tracked without software failure. A triangular wave input was fed into the galvo scanner of the model eye and amplitude was increased (velocity increased) to the point where the tracking began to fail. These failures occur because, at high velocities, the shear of the features within a strip causes the height of the normalized cross-correlation with the reference frame to go below the 0.3 threshold level (see Subsection 2.2). The corresponding threshold velocity was found to be 1761 pixels/sec. Since the velocity of motion in pixels is inversely proportional to the field size, the velocity threshold in degrees per second will depend on field size. Equation (1) establishes this relationship:

$$VelocityThreshold = 3.44 \times FieldSize, \tag{1}$$

where *VelocityThreshold* is the maximum trackable velocity in degrees per second and *FieldSize* is the TSLO field size in degrees. According to this equation, a field size of just over 5.2° would be required in order to correct for the median microsaccade velocity of ~18 °/s [12,13]. For this metric, the larger field size of the TSLO over that previously reported with the AOSLO with a maximum field size of ~2°, offers a performance advantage.

3.3. Stimulus accuracy

The ability to measure eye motion and generate a stabilized video does not directly indicate the accuracy with which a stimulus can be projected onto the retina. While eye motion measurements can be reported at 960 Hz, the delivery of the stimulus involves a prediction and, given the manner in which the stimulus is delivered in the TSLO, some time is required to arm the laser to deliver this stimulus (see Subsection 2.2). We call this time the *latency*. The algorithm is written so that this prediction is completed prior to the beam scanning over the target. Any eye motion that occurs between the prediction and the stimulus delivery results directly in stimulus delivery errors. Therefore, it is important to keep this latency as small as possible.

To compute the latency, fixed frequencies of sinusoidal motion (i.e. 30 Hz) were input into the model eye and the error in stimulus placement was directly observed as relative motion between the dark stimulus and the underlying image. By looking at retinal image strips within the same frame that were recorded prior to the delivery of the stimulus, we could determine the exact location where the motion of the retina was in phase with the stimulus. This location indicated the time point at which the prediction was made, establishing the latency of the stimulus delivery. The stimulus was observed to move in phase with the retinal locations that were 43 pixels away. Given the actual line scanning frequency of 15.74 kHz, these 43 pixels correspond to a latency of 2.73 msec, which is within the expected range given in Fig. 5.

To measure stimulus delivery performace on a human eye, a black circle of 24 pixels in diameter was used as the stimulus. Multiple videos of the second subject were taken at roughly 600 frames each (except for one video with 300 frames). Figure 8 shows a registered sum of 300 stabilized frames from a movie where the stimulus was targeting a retinal location. The sharp stimulus delivery in the average image attests to the tracking accuracy of the system. The video entitled "Stimulus accuracy" in Fig. 9 shows a movie sequence where each frame of the video is cropped around the stimulus. The motion of the retina behind the stimulus is small, but readily apparent. The relative motion between the stimulus and the retinal patch between frames revealed the accuracy of the stimulus delivery [14]. Based on 2100 frames from four movies, stimulus delivery failure occurred in 53 frames, or 2.5% of the frames. As described in Subsection 2.2, failures occur whenever the normalized cross-correlation peak of the image strip used to make the prediction of the target location does not exceed 0.3. The frames in which stimulus delivery did not occur were eliminated from motion

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

error calculations. The standard deviation of motion error in the x direction was 0.65 arcminutes and in the y direction was 0.67 arcminutes. This gives an average standard deviation of motion of 0.66 arcminutes, or roughly the size of an individual cone photoreceptor, for stimulus accuracy.

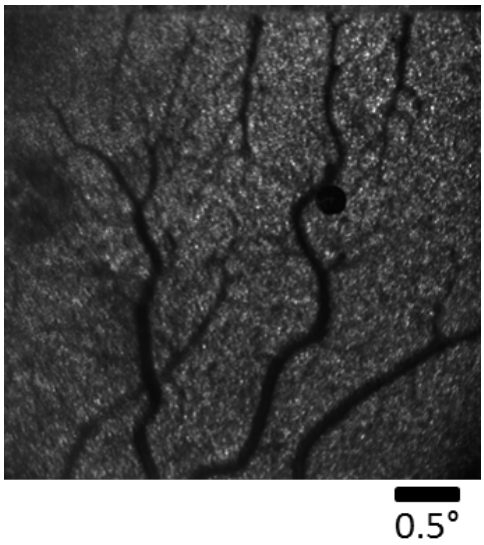

Fig. 8. Registered sum of 300 frames from a movie sequence (Media 1) where the stimulus was tracking a targeted retinal location.

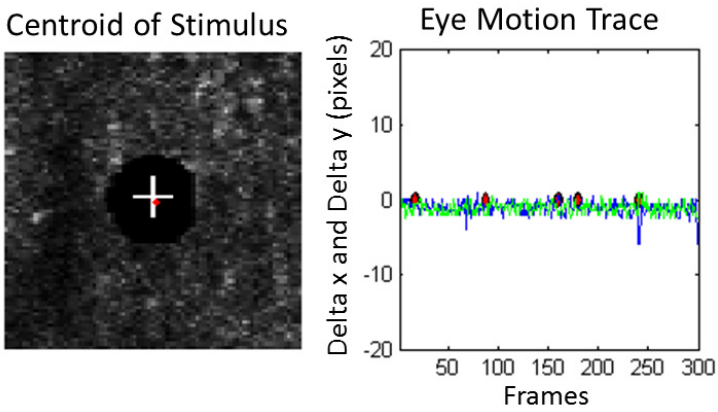

Fig. 9. Left, a single frame from the 300 frame movie entitled "Stimulus accuracy" showing the centroid of the circle stimulus superimposed upon the retina (Media 2). Right, motion error, including microsaccades, for a 300 frame movie. The red diamonds in the graph and media file indicate the frames in which the stimulus delivery failed.

3.4. System resolution

Although this system was designed to be of a wider field size than the typical 1-3 degree field of view of the AOSLO, the field of view may operate anywhere from 1 to 5.5°. By adjusting the angles of each of the scanners, the field size can be made smaller in order to determine the finest structures resolvable. With a 2° field of view, photoreceptor structure starts to become visible outside of the fovea region, with a clear photoreceptor mosiac seen at 4° and beyond without the use of adaptive optics (Fig. 10).

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

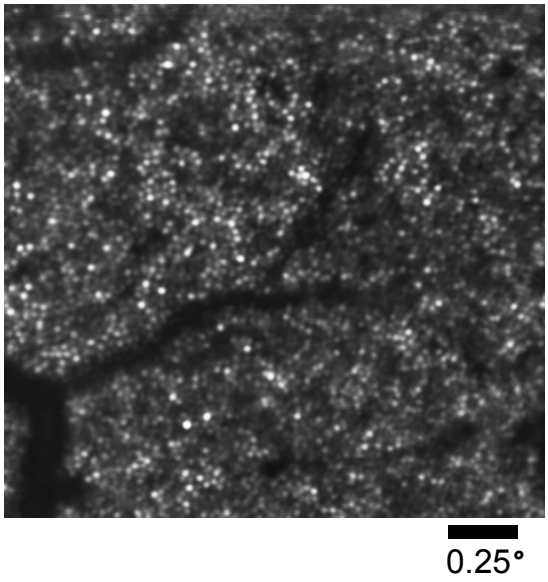

Fig. 10. Image of the cone photoreceptor mosaic of an emmetropic eye at 4° eccentricity from the fovea. The image is an average of 300 stabilized frames from a single 10-second stabilized video. The bright white circles are individual cone photoreceptors, while the dark shadows extending from the lower left corner into the center of the image are blood vessels.

4. Discussion

The TSLO is a robust, high speed, image-based retinal eye tracker that is adaptable to both clinical and research settings. The system may serve as a stand-alone system for recording and stabilizing retinal movies in real-time, as well as providing targeted stimulus delivery for psychophysical experiments. A potential application of the TSLO is that it can also be coupled with a variety of other systems that are in need of accurate eye tracking. These systems include OCT, AOSLO, mfERG and laser guided surgery. The use of the TSLO coupled with these technologies will render high fidelity images without excess motion artifacts, as well as provide an unambiguous record of stimulus delivery onto the retina.

One important question to ask is how does the TSLO compare with other current tracking technologies? Table 1 displays the tracking accuracy, latency and stabilization accuracy of various tracking technologies and methods. The TSLO fares well compared to other tracking methods, with only the AOSLO and the optical lever showing greater accuracy. However, it is important to keep in mind that the tracking capabilities of the TSLO are an extension of those first reported for the AOSLO. The greater accuracy of the AOSLO does come at a cost. The smaller field size of the AOSLO has two main consequences. First, the same lateral motions in a small-field system cause a greater loss of overlap with the reference frame, resulting in a larger number of tracking and stimulus delivery failures due to sub-threshold cross-correlation peak values. Second, the threshold velocity is lower for the smaller field sizes. Additionally, the AOSLO stimuli are limited in their extent due to the small scanning raster of the system (1°- 3°). Therefore, while the increase in field size of the TSLO has lower tracking accuracy, the payoff is fewer software failures caused by larger eye motions, which leads to a more robust performance. In terms of the optical lever method for stimulus control, slippage of the contact lens causes ambiguity in stimulus placement, leaving an uncontrollable amount of retinal movement directly affecting stimulus placement.

While it is clearly shown that the TSLO has many advantages over other tracking technologies, it is important to also understand its limitations as well. First, the reference frame itself will have distortions due to eye motion. Since each frame is built up pixel by pixel, each frame in a movie has unique distortions. Once a reference frame is selected, every

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

Table 1. Comparison of the TSLO to other current eye tracking modalities [4]

Method	Tracking method	Tracking accuracy	Latency	Stabilization accuracy	Comments
TSLO	Retinal image tracking	0.2 arcmin	2.73 msec	0.66 arcmin	Gaze contingent stimulus projection
AOSLO [6]	Retinal image tracking	<0.1 arcmin	3 msec	0.15 arcmin	Gaze contingent stimulus projection. The stimulus is corrected with AO and can be as compact as a single cone.
Optical lever	Direct optical coupling	0.05 arcmin [15]	0 (optical)	0.38 arcmin [16]	Stimulus is very precise but contact lens slippage will cause uncontrollable and unmonitorable shifts in stimulus position
Dual Purkinje (dPi) Eye Tracker with optical deflector [17]	Purkinje reflexes from cornea and lens	~1 arcmin [18]	6 msec	~1 arcmin (error is dominated by tracking accuracy)	
EyeRis™[a] [19]	dPi[b]	~1 arcmin	5-10 msec	~1 arcmin (tracking limited) [20]	Gaze contingent display
MP1 (Nidek, Japan)	Retinal image feature tracking	4.9 arcmin [21]	2.4 msec	Not reported	Gaze contingent display for single stimulus presentations (clinical visual threshold measurements)
Physical Sciences Inc [22].	Retinal feature tracking	3 arcmin	< 1 msec	3 minutes (tracking limited)	Used optical hardware to maintain a scanning raster on its intended retinal location to facilitate high fidelity line scanning ophthalmoscope imaging.
Heidelberg Spectralis OCT(Heidleberg, Germany)	Retinal image feature tracking	Not reported	Not reported	Not reported	Used optical hardware to maintain an OCT B-scan at its intended retinal location to facilitate scan averaging and high fidelity volume imaging.
Eyelink II[c] (250 and 500 Hz)	Pupil only (500 Hz) Pupil and CR (250 Hz)	<30 arcmin	<3 msec	Not reported	Monocular and binocular pupil tracking with a head mounted video based eye tracking system.

[a]This technique falls into a broad class of eye trackers coupled with gaze contingent displays. The EyeRIS system here has the best reported performance of any of the systems we found.
[b]Any tracking method can be used for this type of system but results from dPi system are reported since they provide the best results.
[c]There are many head mounted video based tracking systems used for psychophysical experiments, but the Eyelink II was the most commonly reported system (28 out of 31, or 90% usage) in experiments for measuring human microsaccades in a recent review [23].

subsequent frame will then be stabilized against it. The selection of a reference frame is done manually, using a button press directly in the software interface. If the reference frame is selected during a large microsaccade or blink, the video will not be able to stabilize properly

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA

in real-time, revealing a warped stabilized video. In these situations, the operator has immediate visual feedback on the choice of a reference frame and can always reselect whenever necessary during an imaging session. Second, vertical shifts of the eye can cause a loss of retinal information for eye tracking. Currently, a single frame is used as the reference frame to compute eye motion. If the eye moves vertically, there will be strips that do not overlap with the reference. The TSLO does offer a larger FOV than the AOSLO (currently operating at 5.5° compared to the typical AOSLO usage of 1-2°), which allows one to accurately capture more retinal structure. While this proves highly beneficial for horizontal motion, it still produces error in the vertical direction. Lastly, only eye motions that cause horizontal and vertical displacements of the retina are computed. Rotations of the eyeball about its optical axis, or torsion, are not corrected for properly. The correction of torsion is possible, but the extra computation time needed for its contribution outweighs the benefits of correcting the motion in real-time. It should be noted that even though these errors add artifacts to the eye motion trace, they do not preclude accurate placement of a stimulus at the targeted location.

5. Conclusion

The TSLO is a high speed, robust eye tracking system that can provide high resolution retinal images as well as targeted stimulus delivery. It has been shown that the FPGA solution, with FFT-based cross correlation algorithms, can be translated to traditional scanning laser ophthalmoscope technology. The use of this technology will provide a more compact, robust and cost-effective solution for the study of the retina and fixational eye movements. Using smaller fields of view outside of the foveal region showcases the system's single cell resolution, with individual cone photoreceptors clearly visible without the use of adaptive optics.

Acknowledgments

This research was supported by grants from the Macula Vision Research Foundation (A. R., C. K. S.), the National Institutes of Health EY014735 (A. R., D. W. A., Q. Y.), T32EY007043 (C. K. S.), Stichting Wetenschappelijk Onderzoek Oogziekenhuis Prof. Dr. H.J. Flieringa (SWOO) (J. F. D. B.) and the Combined Ophthalmic Research Rotterdam (CORR) (J. F. D. B.) . Some of the research was performed at the Rotterdam Ophthalmic Institute with assistance from Kari Vienola, Boy Braaf and Koenraad Vermeer. Alfredo Dubra provided helpful conversations regarding system design for limiting astigmatism. Scott Stevenson and Girish Kumar provided the offline motion analysis software. This paper was presented as a poster at the 2012 ARVO Annual Meeting, Fort Lauderdale FL, May 2012.

#173117 - $15.00 USD Received 23 Jul 2012; rev. 17 Sep 2012; accepted 17 Sep 2012; published 19 Sep 2012
(C) 2012 OSA