




Strip-based digital image registration for distortion minimization and robust eye motion measurement from scanned ophthalmic imaging systems

MIN ZHANG,^{1,5,6} ELENA GOFAS-SALAS,^{1,5}  BIANCA T. LEONARD,¹ YUHUA RUI,^{1,2} VALERIE C. SNYDER,¹ HOPE M. REECHER,¹ PEDRO MECÊ,¹  AND ETHAN A. ROSSI^{1,3,4,7} 

¹Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA

²Eye center of Xiangya Hospital, Central South University; Hunan Key Laboratory of Ophthalmology; Changsha, Hunan 410008, China

³Department of Bioengineering, University of Pittsburgh Swanson School of Engineering, Pittsburgh, PA 15261, USA

⁴McGowan Institute for Regenerative Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

⁵Denotes that each of these authors contributed equally to this work

⁶miz62@pitt.edu

⁷rossiea@pitt.edu

Abstract: Retinal image-based eye motion measurement from scanned ophthalmic imaging systems, such as scanning laser ophthalmoscopy, has allowed for precise real-time eye tracking at sub-micron resolution. However, the constraints of real-time tracking result in a high error tolerance that is detrimental for some eye motion measurement and imaging applications. We show here that eye motion can be extracted from image sequences when these constraints are lifted, and all data is available at the time of registration. Our approach identifies and discards distorted frames, detects coarse motion to generate a synthetic reference frame and then uses it for fine scale motion tracking with improved sensitivity over a larger area. We demonstrate its application here to tracking scanning laser ophthalmoscopy (TSLO) and adaptive optics scanning light ophthalmoscopy (AOSLO), and show that it can successfully capture most of the eye motion across each image sequence, leaving only between 0.1-3.4% of non-blink frames untracked, while simultaneously minimizing image distortions induced from eye motion. These improvements will facilitate precise measurement of fixational eye movements (FEMs) in TSLO and longitudinal tracking of individual cells in AOSLO.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Fixational eye movements (FEMs) are an essential aspect of normal human vision that keep the eyes in constant motion [1]. FEMs are physiologically important and serve several useful purposes [1–3]. Though FEMs have canonically been described as consisting of three classes of motion: microsaccades, drift and tremor [4], recent evidence suggests that any tremor in the eye during fixation is likely to be extremely small and thus inconsequential for vision [5]. FEMs are a source of image distortions in ophthalmic instruments that scan an imaging beam across the retina of awake human observers, such as scanning laser ophthalmoscopy (SLO) and optical coherence tomography (OCT). Scanned systems typically scan the retina in a raster pattern implemented with a relatively slow scanner at the frame rate and a faster scanner at the line rate. SLO line rates are often on the order of ~10–14kHz, so each line is usually considered to be free from eye-motion based distortions. However, intra-frame image distortions arise because eye

movements can be much faster than the relatively slow frame rate of scanned ophthalmic imaging systems that typically operate at video rates ($\sim 24\text{--}30\text{ Hz}$). Fortunately, when image sequences are acquired on scanned systems, the motion of the eye is encoded into the image sequence, permitting the motion to be recovered through the application of appropriate techniques [6,7].

Several methods have been developed to recover eye motion from retinal images acquired with scanned ophthalmic systems, such as SLO [7–9] and its higher-resolution implementation in adaptive optics scanning light ophthalmoscopy (AOSLO) [6,10]. Recovery of eye motion for image registration is essential in ophthalmic imaging so that several images from a sequence can be integrated or averaged to generate a high signal to noise ratio (SNR) image of the retina from a sequence of low SNR images. Scanned ophthalmic image registration algorithms have often used a cross-correlation method to compute the offset between a reference image and each image (or sub-image) in an image sequence. To increase the temporal rate of motion measurement, reduce computational cost, and achieve superior results for image registration, most implementations now divide each frame into multiple image strips with a height of several scan lines [11].

Strip-based image registration methods that track motion based on using a single imaging frame as a reference image have successfully achieved high performance at real-time data rates and enabled previously unattainable engineering and scientific aims, such as real-time optical stabilization [11,12] and single-cell psychophysics in AOSLO [13,14]. However, this high performance for real-time applications comes at the cost of a high error tolerance and a reduced sensitivity to large amplitude motion. A high error tolerance results in either dropped strips or poor registration matches. A high error tolerance is acceptable (and sometimes advantageous) for some retinal imaging and most psychophysical applications. Psychophysical applications often tend to consist of hundreds or thousands of trials, so can usually simply just exclude those trials from analysis when the algorithm dropped a strip or gave a bad match. Poor-matches can occur when the image is of poor quality due to optical factors, so the dropping of these poor-quality strips can be advantageous when the objective is to build a registered and averaged image from only the highest quality image strips (e.g. confocal AOSLO). However, the same error-tolerance is unacceptable when the objective is precise eye motion measurement (e.g. to quantify fixational eye movements for tens of seconds or up to minutes), or in cases with low photon flux (e.g. in autofluorescence or non-confocal AOSLO).

A drawback to the single reference frame approach is that it is insensitive to motion that moves the field of view outside the area imaged in the reference frame; this is sometimes referred to as a ‘frame-out’ error. We and our colleagues previously demonstrated methods to increase sensitivity to large amplitude motion for eye tracking in AOSLO using hybrid imaging approaches [11,15]. These add an additional wide-field scanned system to track large amplitude motion in real-time at a coarse scale to drive a tip/tilt mirror to stabilize the small field of view AOSLO enough to prevent most ‘frame-out’ errors. However, these solutions increase the complexity of the imaging system and the data acquisition and image processing pipelines substantially.

Finally, previous approaches that used only a single reference frame do not mitigate or remove the intraframe distortions present within the reference frame but rather encode the reference frame distortion into all the images in the registered image sequence. Though within-frame distortions introduced from the reference frame are usually small (on the order of several microns), they have historically made it extremely difficult to track individual cells longitudinally in AOSLO [16]. Some methods have been proposed to correct intraframe distortions in the reference frame, such as using lag-bias, however these methods work only when the eye movements are stochastic and radially symmetrical, or the predominate drift is compensated with corrective microsaccades. Additionally, they are more likely to fail for large amplitude of eye movement [17,18].

Here we implement a new approach that solves these problems for strip-based registration and demonstrate its applications in TSLO and AOSLO. Free from the constraints imposed by real-time eye-tracking, we devised a more robust approach that achieves our goals for a technique

that: 1) captures the precise motion of nearly all the images in each sequence for eye motion measurement and light starved imaging applications; 2) is sensitive to motion larger than the field of view of a single frame; and 3) reconstructs the spatial arrangement between image features consistently and accurately. These goals are achieved through the implementation of several novel steps compared to previously published approaches, including: 1) a pre-processing step to achieve robust eye motion measurements; 2) a large amplitude motion detection procedure that involves evaluating sub-images to detect motion outside the bounds of a single image frame and 3) the generation of a synthetic reference frame to mitigate the within-frame distortions that are present in every single individual frame acquired from scanned ophthalmic imaging systems and every image registered using a single reference frame approach.

We show here applications of the technique for measuring FEMs with TSLO and for image registration in AOSLO and compare the results of this method to the real-time method of Yang et al. [11]. Our approach was able to track around 99% of all image strips, on average, after excluding blinks and even in the cases of subjectively lower quality data for both AOSLO and TSLO. We also demonstrate that image distortions are substantially minimized with our new technique.

2. Methods

2.1. Participants

All experiments were approved by the University of Pittsburgh Institutional Review Board and adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants following an explanation of experimental procedures and risks both verbally and in writing. Participants ranged in age from 14 to 59 (average: 20; female: 20; male: 21) and were compensated for their participation. To ensure that imaging was safe, all light levels were kept below the limits imposed by the latest ANSI standard for safe use of lasers [19].

2.2. Data sources

We evaluated our approach using sixty image sequences that had been previously acquired on our TSLO and AOSLO systems for ongoing studies [20,21]. Since we were interested in evaluating performance on a range of image qualities, we selected ten image sequences from each system in three general levels of image quality. Images were graded subjectively by two of the authors (MZ and EG) to be of either low, medium, or high quality; both graders had to agree on the grading for the image sequence to be included in the evaluation dataset. Examples of each quality level for each device are shown in Supplementary Fig. S1 in [Supplement 1](#).

2.2.1. Tracking scanning laser ophthalmoscopy (TSLO)

The tracking scanning laser ophthalmoscope (C. Light Technologies, Inc., Berkeley, CA) has been described in detail [12]. An 840 nm (50 nm bandwidth) super luminescent diode (SLD) provided illumination over a field size of $5^{\circ} \times 5^{\circ}$. Participants were imaged sitting in a chin rest that was stabilized with temple pads. Image sequences (512×512 pixels) were acquired monocularly, from the left eye, without dilation, at 30 Hz for 30 seconds.

2.2.2. Adaptive optics scanning laser ophthalmoscopy (AOSLO)

The Pittsburgh adaptive optics scanning laser ophthalmoscope has been described in detail elsewhere [21]; image sequences were used from the confocal imaging channel only. A 795 nm (FWHM = 15 nm) SLD provided illumination over a field size of $1.5^{\circ} \times 1.5^{\circ}$. Image sequences (512×496 pixels) were acquired monocularly, from either the left or right eye, with dilation, at 29 Hz for 10–180 seconds.

2.3. Algorithm workflow

The algorithm workflow, implemented in MATLAB (R2018a; The MathWorks Inc., Natick, MA), is outlined in Fig. 1. It consists of three main steps: 1) pre-processing and detection of highly distorted frames; 2) coarse registration, large motion detection and synthetic reference frame generation; and 3) fine strip-level registration. The overall strategy follows the general framework described by Stevenson and Roorda [6] and that we and our colleagues have built upon [10–12]. However, we have implemented a new end-to-end approach here to achieve the goals outlined previously. Each step is described in detail in the sections below. Our workflow begins with blink detection followed by a pre-processing stage.

2.3.1. Blink detection

Blinks were detected using a rudimentary intensity threshold to $\log\text{Mean}(i)$, where $i \in [1, 2, \dots, m]$ is the index of each image and m is the total number of images in the sequence. $\log\text{Mean}(i)$ is computed by normalizing the mean intensity of each image (i) to the interval of $[0, 1]$, followed by applying the common logarithm on the normalized mean. Note that NaN will be used if the common logarithm of 0 is encountered and hence will not be considered in the threshold calculation. The threshold was defined as: $\text{blinkth} = \text{minimum}(\log\text{Mean}(i))/3.5$. Images with $\log\text{Mean} \leq \text{blinkth}$ are marked as blinks and excluded from further processing. It should be noted that this method may erroneously detect blinks for motion traces that do not contain a blink and have uniform intensity across all frames.

2.3.2. Pre-processing

Pre-processing reduces noise, improves contrast, and minimizes large gradients in intensity across the field of view. Non-uniform image intensity across the field of view can result from highly scattering structures in the normal retina, such as the foveal reflex; they can often also arise anywhere in the retina in disease states such as age-related macular degeneration. A strong foveal reflex is often seen in younger healthy eyes and was observed in much of our TSLO data. Pre-processing started with Gaussian filtering, implemented with MATLAB's built-in `imgaussfilt` function (with $\sigma = 20$), to remove high-frequency noise. This was followed by contrast-limited adaptive histogram equalization (CLAHE) using the `adapthisteq` function (with `ClipLimit = 0.05`). CLAHE serves to effectively improve local image contrast. These pre-processing steps were implemented with the same fixed parameters for both TSLO and AOSLO and improved the detection of the normalized cross-correlation (NCC) peak that is used for subsequent processing stages (see Suppl. Fig. S2 in Supplement 1). It should be noted that these come at little computational cost, representing a negligible proportion of the overall processing time for each image sequence ($\sim 0.4\%$).

Highly distorted images were identified by computing the NCC for each pair of consecutive images in each sequence, similar to Salmon et al. [22]. All NCC computations were performed with MATLAB's built-in `normxcorr2` function, using the CUDA implementation with elements of the Parallel Computing Toolbox. Registration was carried out on machines equipped with Nvidia GPUs (GTX 1080). Since highly distorted images have little or no overlapping features with both the previous and consecutive frame in the sequence, the peak of the NCC matrix computed between these frames is low. We identified candidate distortion frames based on this principle by applying a simple threshold based on the statistics of the image sequence. When the peak in the NCC matrix between the previous frame and consecutive frame were both less than the threshold, we considered these frames to be distortion frames. The threshold was defined as $\text{NCC} < \mu - 0.8\delta$, where μ and δ are the average and SD of the NCC peak for the entire image sequence. Distortion frames were excluded from further analysis.

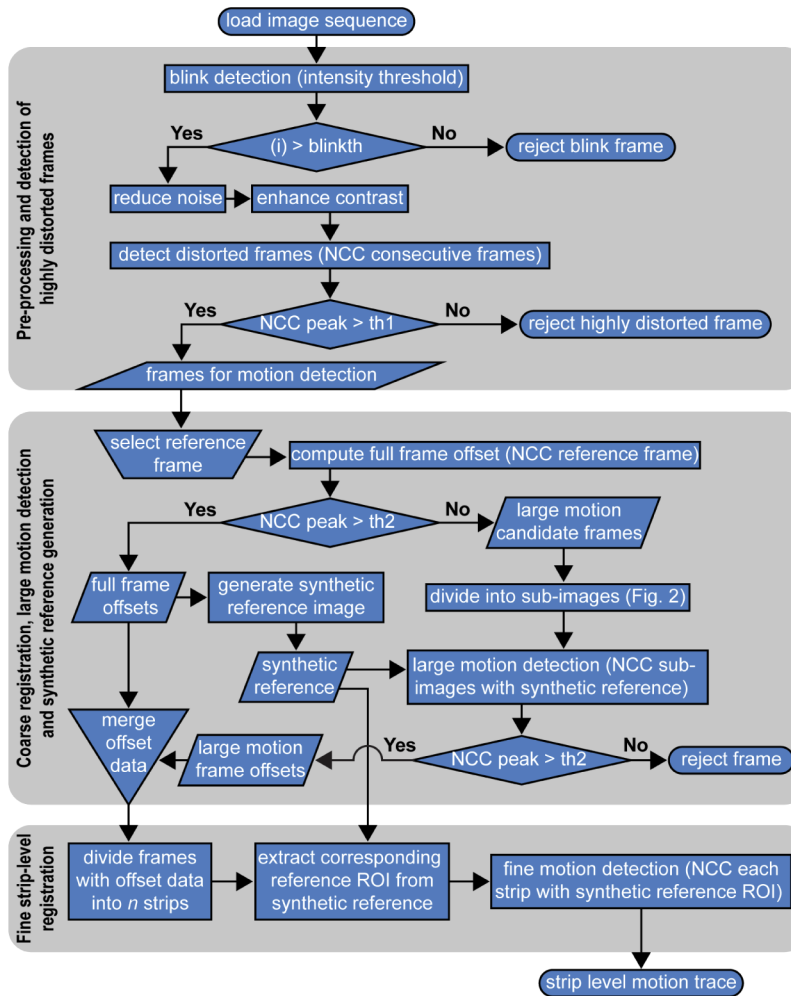


Fig. 1. Algorithm workflow. An image sequence is loaded to begin the pre-processing stage (top grey rectangle). Frames with intensity (i) below the blink threshold (blinkth) are discarded while those above undergo noise reduction and contrast enhancement. This is followed by the detection of distortion frames (NCC on consecutive frames). Frames below threshold 1 (th_1) are rejected as highly distorted frames. Frames above th_1 are passed to the next stage for coarse registration, large motion detection and synthetic reference frame generation (middle grey rectangle). A reference frame is selected manually, followed by a full frame offset NCC calculation using the manually selected reference. Frame offsets detected with an NCC peak greater than threshold 2 (th_2) are applied to generate the synthetic reference image. Those frames below th_2 are divided into sub-images (see Fig. 2) for large motion detection (NCC between sub-images and synthetic reference). Those sub-images with an NCC peak below th_2 are discarded while those above th_2 are used to detect the large motion offsets that are merged with the offset data used to create the synthetic reference. The merged offset data is then passed to the final stage of processing, fine strip-level registration (bottom grey rectangle). Strip-level registration is computed by dividing each frame with offset data into n strips. The full frame offset data is used to determine the position of the corresponding reference ROI on the synthetic reference. Fine-scale motion is detected by computing the NCC between the reference ROI and the strip (see Fig. 3). A strip-level motion trace is then output.

2.3.3. Coarse alignment and identification of candidate 'large motion' frames

This step computes coarse frame-to-frame offsets and synthesizes a composite reference frame for fine motion extraction. This begins with the selection of a reference frame by the experimenter (e.g. Fig. 2(a)). The criteria we used for selecting a reference frame was that it should be of the average image quality of the rest of the sequence (i.e. not blurry or of low quality) and free from visible distortions. It should be noted that even though most individual frames do not contain visible distortions (such as visible compression or shearing artifacts; see Suppl. Fig. S3 in Supplement 1 for examples), all frames do contain distortions due to eye motion during the frame. This is due to the slow frame rates of the imaging systems; since it takes 33 ms to acquire a single frame at 30 fps there will necessarily be some motion during this interval. These insidious distortions prevent us from achieving distortion free imaging with a single reference image and that is why we follow this step with the generation of a synthetic reference frame to minimize these within-frame distortions that are present in every single frame of every image sequence.

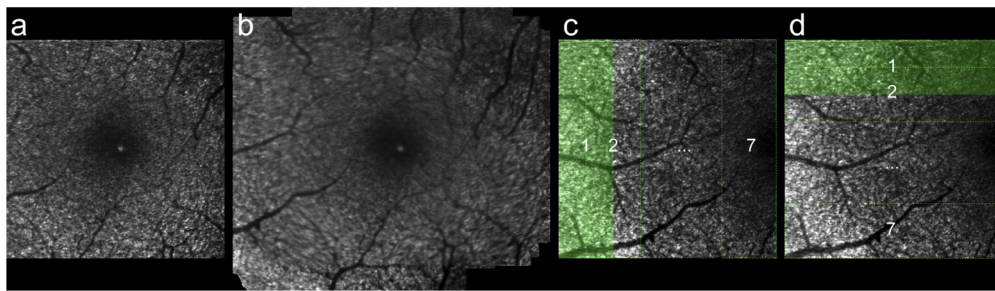


Fig. 2. Coarse alignment for reference synthesis and detection of 'large motion' candidates. An image from the sequence (TSLO in this example) is chosen manually to serve as the reference for coarse alignment (a). Coarsely aligned frames are averaged to generate a synthetic reference frame (b). Large candidate motion frames, (c) and (d), are divided into 7 strips both vertically (c) and horizontally (d) and cross-correlated with the synthetic reference frame (b) to capture large amplitude motion.

The full-frame NCC is then computed between each of the frames and the synthetic reference frame for motion detection (see Fig. 1), as we and others have described previously [11–13,23]. In practice, some image frames may have a small overlap with the reference frame due to relatively large eye motion (e.g. Fig. 2(c)), resulting in a reduced NCC peak. To maximize the chances that these frames could be captured in our analysis, we set a second threshold at $\mu - 0.6\delta$ to extract these frames as a group of candidates of 'large motion' frames; large motion frame candidates were then reserved for additional processing steps (described below).

2.3.4. Generation of a synthetic reference frame and evaluation of large motion candidates

The next step was to synthesize a larger reference frame for the fine motion trace computation. A larger reference frame enables motion that goes beyond the bounds of a single reference frame to be captured. This was generated by simply averaging the registered images from the coarse alignment step (see Fig. 2(b)). Next, we attempted to capture the coarse offsets for the images that we determined to be large eye motion candidates. We did this by dividing each of those frames into seven strips vertically and horizontally (see Fig. 2(c) and 2(d)). Strip size for this procedure was 512×128 pixels or 128×512 pixels, with 64 pixels of overlap. We then calculated the NCC between each strip and the synthetic reference image. The offset of the strip with the highest NCC peak was selected to represent the true offset of the corresponding frame. If the NCC peak of all the strips for that frame was still low ($< \mu - 0.6\delta$), the image was rejected from further analysis.

2.3.5. Fine motion trace extraction

In the previous step we identified and flagged distortion frames, computed frame-to-frame offsets, generated a composite reference frame and evaluated frames with the largest motion. In this step, we divide each image whose coarse motion was successfully captured, into multiple image strips as shown in Fig. 3. The strip size and overlap can be specified by the user; we used 16 strips per image (i.e. 32×512 pixels) herein for a nominal temporal sampling rate of 480 Hz for the motion traces. For each strip, we identified its coarse matching region in the synthetic reference frame from its coarse frame offset. We then extract a larger region of interest (ROI) reference strip by extending the region one strip height above and one strip below the coarse matching region (Fig. 3(a)). We then use this ROI reference strip for the NCC computation for that strip rather than the whole synthetic reference frame. This approach reduces the computation cost and has the potential to increase accuracy, since the best matching location is usually within the ROI reference strip.

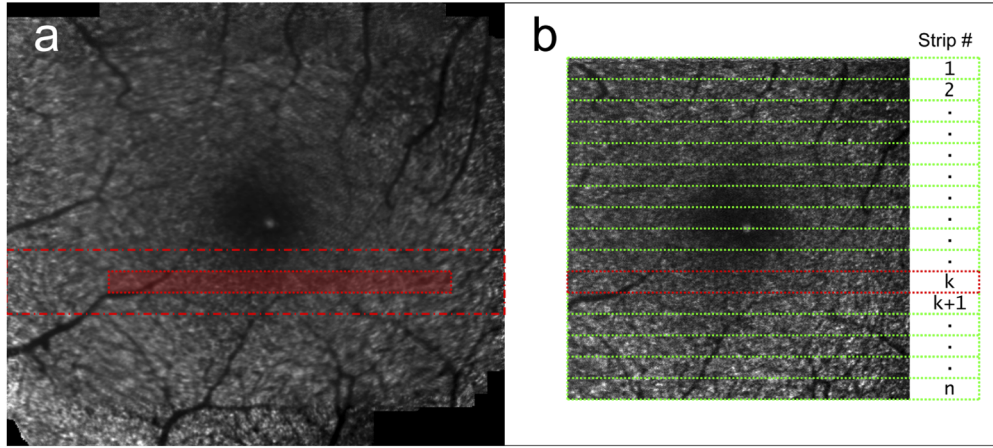


Fig. 3. Strip-level motion trace extraction. The synthetic reference image (a) serves as the source of reference ROIs for fine strip-level motion extraction. Each target frame (b) is broken up into strips (e.g. strip k , outlined in red in b) and its probable position on the synthetic reference (shaded area in a) is estimated from its coarse frame offset. A region of interest (ROI) reference area (denoted by the dashed red rectangle in (a)) is then defined and fine motion is extracted by computing the NCC between each strip and its corresponding reference ROI area. Images are from TSLO.

For the example shown in Fig. 3, if $Y_{\text{frm_offset}}(i)$ and $X_{\text{frm_offset}}(i)$ denote the frame offset of the raw target frame (Fig. 3(b)), and the region of its k th strip is $[x_k, y_k, w, h]$, where (x_k, y_k) is the starting position of the strip, w is the width and h is the height of the strip (in our case $w = 512$ pixels, and $h = 32$ pixels), then its matching region in the reference frame (shaded red area in Fig. 3(a)) can be determined by $[x_{\text{ref}}, y_{\text{ref}}, w, h]$, where

$$x_{\text{ref}} = x_k + X_{\text{frm_offset}}(i) \quad (1)$$

$$y_{\text{ref}} = y_k + Y_{\text{frm_offset}}(i) \quad (2)$$

and its ROI reference strip can be determined by $[x_{\text{roi}}, y_{\text{roi}}, w_{\text{roi}}, h_{\text{roi}}]$, where w_{roi} equals to the width of the synthetic reference frame, and

$$x_{\text{roi}} = 1 \quad (3)$$

$$y_{\text{roi}} = y_{\text{ref}} - h \quad (4)$$

$$h_{roi} = 3 * h \quad (5)$$

then we calculate the NCC using this strip and its corresponding ROI reference strip to identify the best matching location, and get its strip offset to represent the eye motion. In those rare occasions when two peaks were identified we used the one closest to the frame offset as the strip offset. We repeat this operation on all the frames to extract the strip offsets for the entire image sequence.

2.4. Performance assessment

2.4.1. Comparison to benchmark

Since no ground-truth of the motion of the eye exists for our datasets, we chose a few different ways to evaluate its performance. One approach we took was to compare our new method to the registration method of Yang et al. [11] that we considered for the purposes of this report to be our gold-standard benchmark for AOSLO. Unfortunately, there does not exist a corresponding method for TSLO and there are fundamental differences between the two techniques that make this an imperfect comparison with several drawbacks (see discussion).

In our first attempted comparisons, we evaluated the results of each algorithm using comparable user-defined variables that govern the strip-wise registration, including strip size, number of strips and strip rejection threshold. As our studies of fixational eye movements required motion traces at a nominal temporal sampling rate of at least ~480 Hz, we developed our approach using 16 strips per frame and a strip height of 32 pixels and used these settings in each algorithm for our initial tests for both AOSLO and TSLO datasets. However, we found that these parameters often caused the benchmark method to perform much more poorly on the AOSLO data than it did with its default parameter set. So, we tested a range of values and settled on using the default settings for AOSLO in the implementation of the offline digital registration version we had access to, that was implemented with a default of 15 strips per frame and a strip height of 64 pixels as this empirically gave the best results. The strip rejection threshold also differed as we used a variable threshold (see above), while the benchmark used a fixed threshold of 0.75.

To compare registration methods, we evaluated several aspects of the results, including: 1) the proportion of successfully registered data; 2) the standard deviation of pixel intensity in the registered image sequences; and 3) the energy of the high spatial frequency information in the final registered and averaged image. We also assessed structural repeatability in the registered and averaged images by comparing the variability in the spatial relationship between image features in the resulting images when different starting reference frames were manually selected.

The proportion of tracked frames was computed by dividing the count of successfully tracked frames by the total number of frames in the sequence after excluding the blinks. It should also be noted that blinks are detected differently in the benchmark algorithm; it uses a cross-correlation threshold to detect blinks [11] rather than the intensity-based approach we use here. For our fixational eye motion measurement work using TSLO, we excluded from analysis discontinuities in motion traces of less than a single frame, so compare the proportion of tracked frames here for both TSLO and AOSLO.

The variation in pixel intensity between the registered and unregistered image sequences was evaluated based on the hypothesis that after registration each pixel remains fixed on the same structure and thus experiences less variability in intensity over time than in the raw data where each pixel continuously sweeps across different structures. To assess this, we computed the standard deviation (SD) of each pixel across time; this was compared between the original and registered image sequences. This was done for cropped regions about the center of 256×256 pixels each of the original and registered image sequences to exclude the margins of the image that may have had few strips contributing to each pixel and to ensure that the same area was compared. In addition, we also evaluated the SD of a registered image sequence that was composed only of frames that were dropped by the benchmark method but successfully tracked by the present

approach. This allowed us to evaluate whether the additional frames tracked by the present approach were registered to the same level as those that were tracked by both.

The image energy of the high spatial frequencies in the image was computed using the registered and averaged images produced from each algorithm through the following steps: Each averaged image was filtered using a high pass filter with a normalized cutoff frequency of 0.02; the spatial variance was computed. This metric informs about the amount of energy in high spatial frequency features and therefore its value will decrease as image blur increases [24]. Finally, we computed the difference between the image energy generated from both registration methods. We plotted this difference divided by the energy of averages registered with our method in order to show the difference in energy as a percentage of the total energy of the averages.

2.4.2. Manual landmark-based performance evaluation

Finally, we also employed a manual method to evaluate whether the algorithm was appropriately registering image features that were easily identifiable by eye in the image sequence. To do this, we randomly picked 20 frames from each image sequence and manually marked the same features that were easy to identify in the images (e.g. vessel crossings). We selected several landmarks for each frame (see Fig. 4). Multiple landmarks were chosen across the field of view to evaluate strip level accuracy across the entire image and to ensure that as the eye motion moved the field of view from frame to frame, that some landmarks would be visible in each image in the sequence. The manual marking was carried out by two independent graders, to allow us to compare the different human graders to one another and to the different registration algorithms. Bland-Altman [25] plots were generated to evaluate the agreement between the different methods.

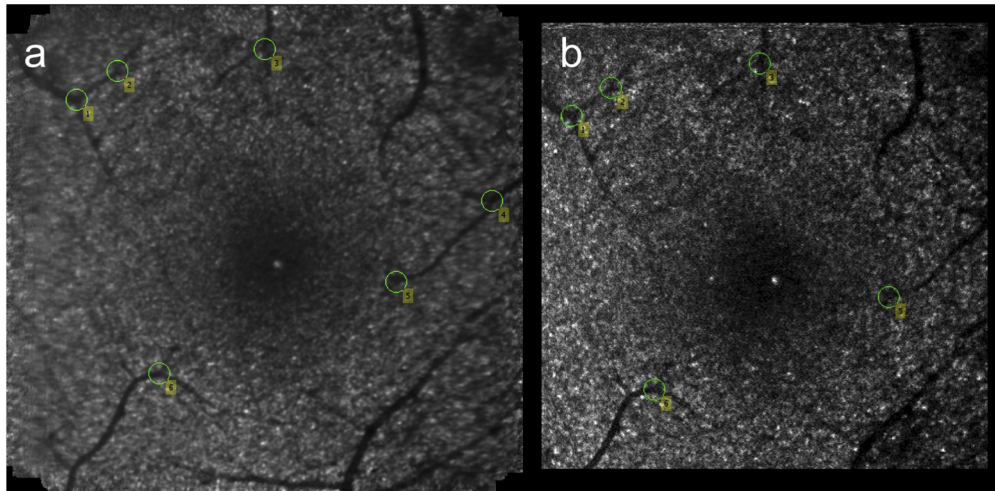


Fig. 4. Manual landmarking. Several vessel landmarks were marked manually in MATLAB. Screenshots from the marking script are shown here with several landmarks marked on the synthetic reference image (a) and then marked in an example target frame (b). This manual landmarking allowed for a direct comparison between the locations of readily identifiable image features marked by manual graders and their positions determined algorithmically. Images are from TSLO.

2.4.3. Structural repeatability across different manually selected reference frames

To evaluate the structural repeatability of the final registered and averaged images, we compared the results of our algorithm to the benchmark when different reference frames are selected manually as the starting reference. For this, we used AOSLO image sequences and generated

registered and averaged images for several starting reference frames. We then co-registered all of the averaged images generated from each algorithm, using sub-pixel registration [26], and compared the spatial arrangement of the features across reference frame for each algorithm. These images were assessed both by visually evaluating animations generated from the registered images and by carefully comparing the positions of individual cone photoreceptors across the different images.

2.5. Results

2.5.1. Comparison to benchmark

Across image quality levels in our TSLO datasets, our method successfully tracked 99.79% of all non-blink frames, leaving only 0.21% of all non-blink frames not successfully tracked. In comparison, the benchmark successfully tracked 68.89% of all non-blink frames across all quality levels, leaving 31.11% of frames untracked, on average. It should be noted that the blink detection methods are different between our method and benchmark, 6.66% of frames were detected as blinks by our method, while 5.21% of frames were detected as blinks by benchmark. Table 1 lists the results for each of the different quality levels across the TSLO datasets and demonstrates that there were not major differences in performance across the range of image quality levels tested for the present algorithm with the proportion of unsuccessfully tracked frames ranged from 0.04–0.43%. In comparison, the benchmark was unable to track between 20.31 and 37.29% of the frames across each quality level.

Table 1. Proportion of tracked frames calculated excluding blink frames^a

System	Image Quality	Proportion (%) of frames	Present method	Benchmark
TSLO	All (n = 30)	Tracked:	99.79%	68.89%
		Not tracked:	0.21%	31.11%
	High (n = 10)	Tracked:	99.96%	64.27%
		Not tracked:	0.04%	35.73%
	Medium (n = 10)	Tracked:	99.57%	79.69%
		Not tracked:	0.43%	20.31%
	Low (n = 10)	Tracked:	99.85%	62.71%
		Not tracked:	0.15%	37.29%
AOSLO	All (n = 30)	Tracked:	97.31%	81.76%
		Not tracked:	2.69%	18.24%
	High (n = 10)	Tracked:	99.62%	96.69%
		Not tracked:	0.38%	3.31%
	Medium (n = 10)	Tracked:	96.61%	61.51%
		Not tracked:	3.39%	38.49%
	Low (n = 10)	Tracked:	97.59%	89.43%
		Not tracked:	2.41%	10.57%

^aThe proportion of tracked frames was similar across the range of image qualities tested for both the TSLO and AOSLO datasets. A larger proportion of frames were not successfully tracked by the benchmark algorithm across all test data.

Across image quality levels in our AOSLO datasets, our method successfully tracked 97.31% of all non-blink frame strips, leaving only 2.69% of all non-blink frame strips not successfully tracked. In comparison, the benchmark successfully tracked 81.76% of all non-blink frames across all quality levels, leaving 18.24% of non-blink frames untracked, on average. Again, the blink detection in our method was different from benchmark: 6.02% of frames were detected

as blinks in our method, in comparison, the benchmark detected 10.43% of frames as blinks. Table 1 lists the results for each of the quality levels across the AOSLO dataset. Again, there were not major differences in performance across the range of the image quality levels tested for the present algorithm; the proportion of unsuccessfully tracked frames ranged from 0.38–3.39%. In comparison, the benchmark was unable to track between 3.31% and 38.49% of the frames across each quality levels. To compare the differences between the results visually, we have provided a side-by-side comparison between a raw AOSLO image sequence and the registered image sequences produced by each method in [Visualization 1](#).

To compare differences between the resulting motion traces, we have plotted a trace from the benchmark overlaid with a trace from the present method for one dimension of a TSLO image sequence in Supplementary Fig. S4 (see [Supplement 1](#)). This comparison shows that both methods detected most of the actual blink frames, that were verified by visually inspecting these frames, as blinks (Suppl. Fig. S4, black frames). However, we also see that there were some differences in the additional frames each method also labelled as blinks. First, we see that our approach always detected slightly longer blink intervals as our intensity-based approach labelled as blink frames some lower intensity frames before and after each shared blink interval. These are the frames that occur during the opening and closing periods of the eye when the eye is still partially opened (Suppl. Fig. S4, dark blue frames). These frames were untracked by the benchmark but not marked as blinks (Suppl. Fig. S4 in [Supplement 1](#), yellow frames at 1 and 17 sec). The benchmark also detected a few frames as blinks that were not blinks (Suppl. Fig. S4 in [Supplement 1](#), cyan frames) due to the cross-correlation threshold-based approach in the benchmark for blink detection that will mark all frames below the threshold as a blink. Some highly saturated frames with larger motion were discarded by both algorithms (Suppl. Fig. S4 in [Supplement 1](#), green frames). Finally, we see that the present method discarded some frames (Suppl. Fig. S4 in [Supplement 1](#), orange frames) that the benchmark did not, including some highly distorted frames. To visually evaluate another TSLO eye trace generated by our approach along with the corresponding image sequence, we have generated the animation shown in [Visualization 2](#) that shows the cumulative motion trace along with the original image sequence.

Figure 5(a) shows the histogram of the SD across time for each pixel in the original data, Fig. 5(b) and Fig. 5(c) show histograms of the SD across time for each pixel in the registered image sequences created for the frames that were successfully registered by both algorithms; there is less variability in intensity over time in the registered image sequences than in the original raw data; in addition, there is very little difference in the distribution of each as is demonstrated by the difference in Fig. 5(e). Similarly, Fig. 5(d) displays the histogram of SD across time for each pixel in the frames that were successfully tracked by our algorithm but that were not tracked by the benchmark; again, the distribution is very similar with the difference histogram in Fig. 5(f) showing very little difference between the two distributions.

2.5.2. Landmark comparison

We compared the manual landmarks made between the two graders and compared the differences between the graders and the algorithms; we also evaluated the agreement between the two algorithms. Bland-Altman plots evaluating the agreement between these different measurements are provided in Supplementary Fig. S5 in [Supplement 1](#). The mean difference between the manual marking by different graders was less than a pixel (Suppl. Fig. S5(a) in [Supplement 1](#); -0.1 px horizontally and -0.3 px vertically). This average difference was smaller than the average difference obtained between the two algorithms that averaged 1.8 px horizontal and 1.5 px vertical (Suppl. Fig. S5(b)). Comparison between the manual graders and the two algorithms are shown in Suppl. Fig. S5(c) and Suppl. Fig. S5(d) in [Supplement 1](#). The average difference between the landmark method and our algorithm was a fraction of a pixel: -0.4 px horizontal and -0.2 px vertical, while there was a larger average difference for the benchmark of 2.2 px horizontal and

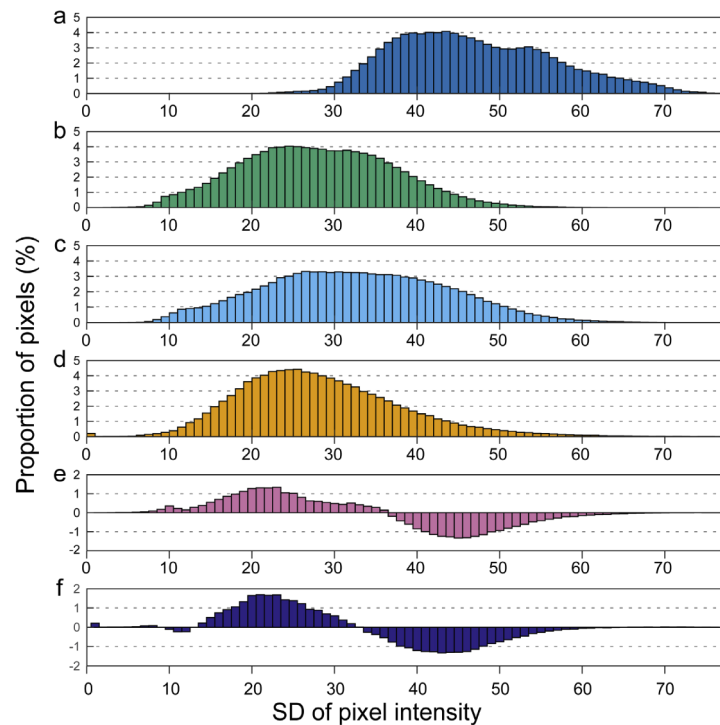


Fig. 5. Standard deviation of pixels across time is reduced similarly for frames successfully tracked by each algorithm. The distribution of pixel SD across time for the (a) original raw data, for the registered data with the (b) current method and (c) the benchmark for frames successfully tracked by each, and (d) for the registered data only successfully tracked by our current method. (e) is the difference between distributions (b) and (c) showing that the distribution was similar for these two methods for frames successfully tracked by each. (f) is the difference between the distributions (c) and (d), showing that there was not a substantial difference in the proportion of pixels successfully tracked between the shared frames (e) or the frames only successfully tracked by the current method (f). These are calculated using the registered image sequences that were constructed using the original raw 8-bit grayscale data (range 0–255).

–1.7 px vertical. Careful inspection of the Bland-Altman plots across the various comparisons shows that the range defined by the 95% confidence limits is greatest for the landmark versus benchmark comparison (Suppl. Fig. S5(b); ~13–14 pixels), while the range observed for the differences between the present method were similar when compared either to the landmark or the benchmark (Suppl. Fig. S5(a) and S5(d); ~11 pixels). The range was smallest when comparing the two graders to one another (Suppl. Fig. S5(d) in [Supplement 1](#); ~8 pixels). We did not observe any trends in the data for increasing (or decreasing) differences with increased averages across the range of comparisons plotted. Taken together, these plots demonstrate that there is reasonable agreement across all measurements compared.

2.5.3. Comparison of high spatial frequency information in registered and averaged images

We computed the high energy spectral band for each of the 30 registered and averaged images obtained in both TSLO and AOSLO. Figure 6 shows the normalized difference between these energy values across the images. For TSLO, there is a trend of increasing difference in energy between the two methods as image quality decreases. The inset shows two examples for two

registered images that fall on different sides of the quality spectrum. For the high quality image sequence there is no discernable difference in the resulting averaged image (cyan arrow and cyan outlined images) while the averaged images computed for the low-quality image sequence shows a much sharper and higher contrast image (green arrow and outlined images). This qualitative result is borne out in the energy metric. For the first example the difference in energy is almost none while for the second it is almost 75%. The same metric applied to the AOSLO image sequences is shown in Fig. 6(b). For this layer (i.e. photoreceptor layer) with this instrument both algorithms have a similar performance as the difference in energy is mostly close to zero except for two exceptions. For image sequence number 16 our algorithm seems to underperform with respect to the benchmark. A region of the averages on the right side of the plot is enlarged showing how the photoreceptors from the image produced by our method are indeed blurrier than the ones on the image obtained with the benchmark.

2.5.4. Structural repeatability

For AOSLO, we also evaluated the structural repeatability of the images by comparing the resulting averaged images when the registration was seeded with different manually selected reference frames. The differences in the spatial position of image features are best appreciated by evaluating the animation shown in [Visualization 3](#). This animation shows the averaged images, co-registered for each method, side-by-side for a representative high quality AOSLO image sequence when 5 different reference frames were manually selected. It should be noted that each reference frame appeared to be free from obvious eye motion distortions and considered to be of equivalent subjective image quality. This animation shows that there is little variation in the spatial arrangement of image features for our method, while the images obtained from the benchmark method show variations in the positions of each cone from image to image. This variation in cell position is evaluated for several cone across these different images in Fig. 7.

Here we show the resulting image for the first reference frame shown in the animation (for reference frame 1) on the left side. The 8 colored squares arrayed vertically across the image denote the location of the zoomed in views of those regions that are shown for each of the five reference frames in the colored rectangles to the right. Within each of these small regions of interest that are shown with 2x magnification to the right we have denoted the position of two of the cones in the first image with colored circles. The location of these cones in the image generated using the first reference frame is overlaid on the other four images to show whether that cone remained in the same location across references or whether it shifted position. Nearly every cone was in the same location for each of the cones evaluated in the images generated from our method. There are some small shifts that can be appreciated in the animation and in a few instances here we see some small variability in cone position across reference frames (e.g. slight shifts for the lower cones in the yellow region and in the light blue region for images 3-5). The variability in cone position was greater for the benchmark method and could result in position differences across the different images on the order of the size of an individual cone (e.g. orange, green, and cyan areas).

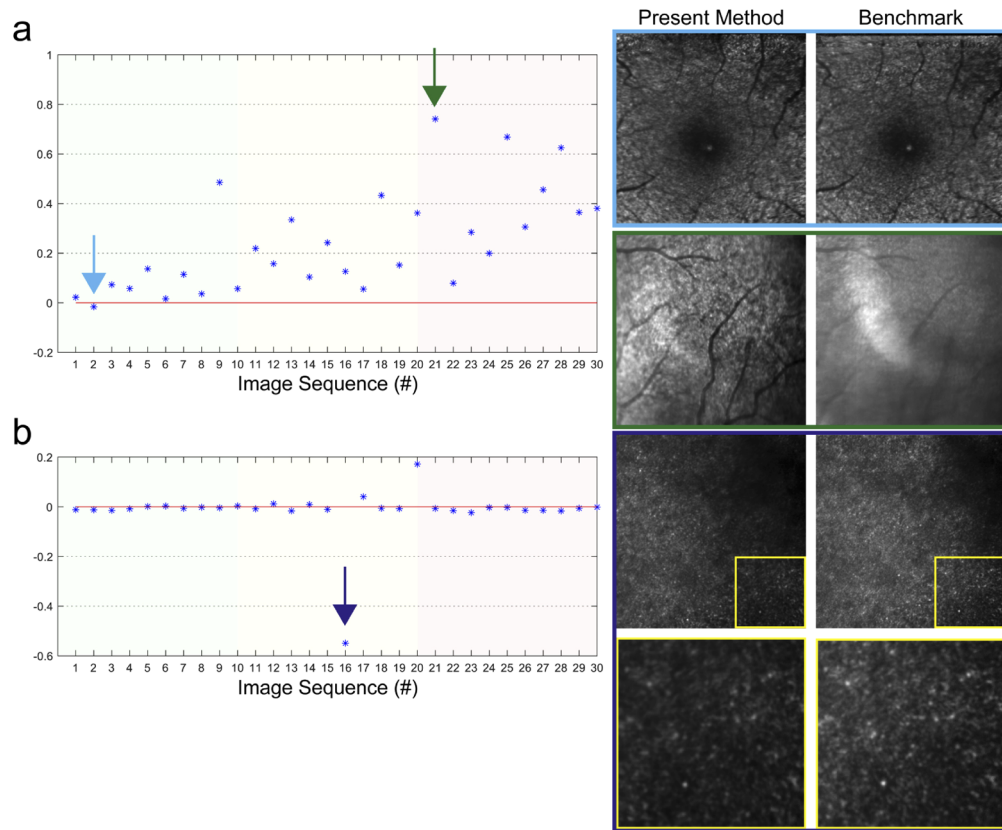


Fig. 6. Normalized difference in energy of high spatial frequencies between our method and the benchmark. TSLO (a) and AOSLO (b) registered and averaged images across all data are compared here. Background color denotes subjective quality of the image sequence (high = green; medium = yellow; red = low). TSLO datasets showed a range of differences, with small differences resulting in negligible differences in subjective image quality (e.g. cyan arrow in (a) and corresponding images traced in same color to the right). There were larger differences for lower quality image sequences and the present method produced images with subjectively higher quality than the benchmark (e.g. dark green arrow in (a) for corresponding averaged images traced in same color to the right). The difference was very small for most AOSLO image sequences aside from one outlier that demonstrated better subjective image quality from the benchmark compared to the present method (purple arrow in (b) and images to right traced in corresponding color). This is appreciated better by examining the zoomed in section (bottom images traced in yellow, location denoted by yellow squares in the larger images above).

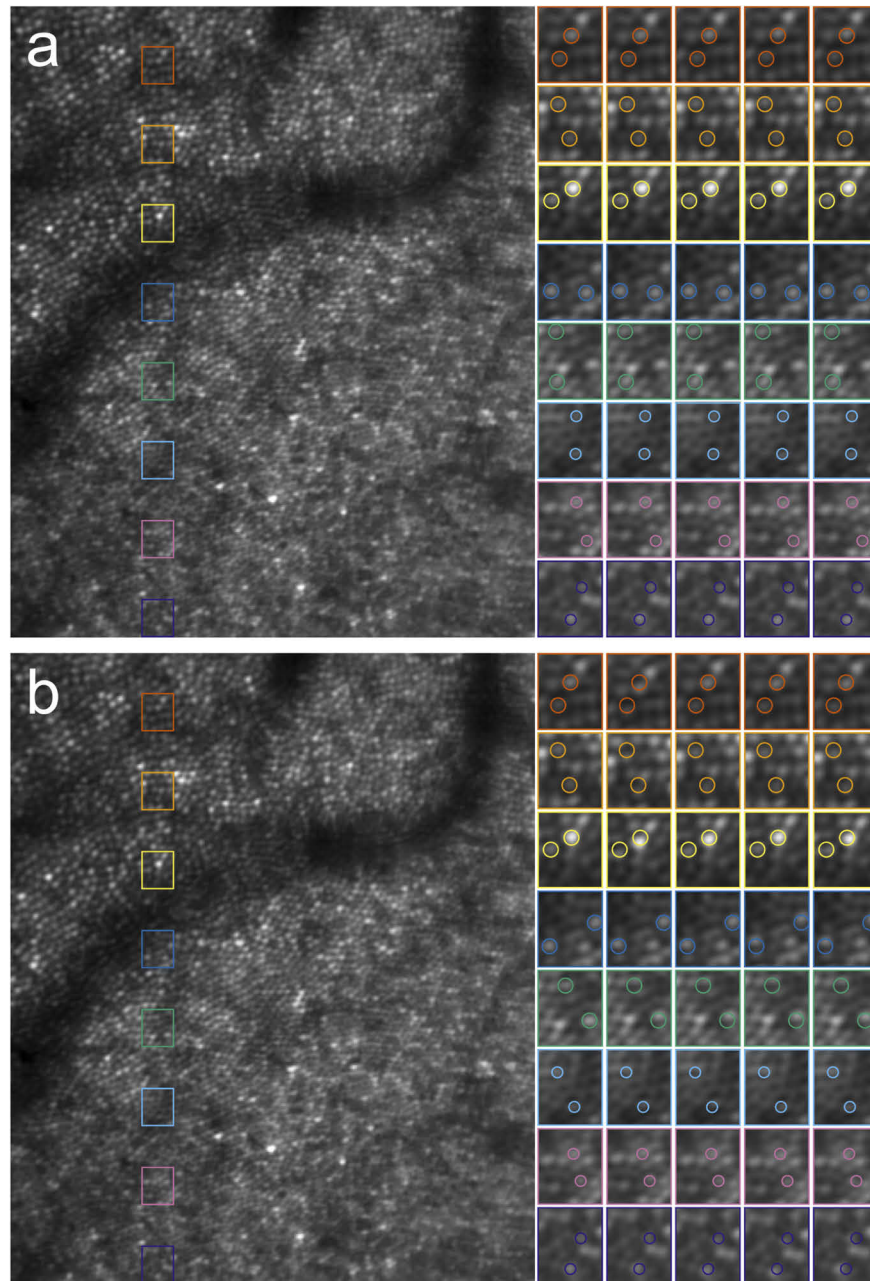


Fig. 7. The same spatial arrangement of cells is seen no matter what reference frame is selected. The present method (a) and benchmark (b) were used to generate five different registered and averaged images from the same AOSLO image sequence when starting with the same five different reference frames. The full averaged image generated when using the first reference frame is shown at left. The overlaid colored rectangles in each denote the positions of the location of the zoomed in locations shown to the right for each of the different reference frames. Two cones were circled in the first image (zoomed images, left column) and then overlaid at the same location on the other images to evaluate cell location repeatability. Nearly all cones for the present method remained in the same location no matter what reference was used. This was not the case for the benchmark, where many cones (e.g. orange, yellow, light blue rectangles) shifted positions depending on the manual reference frame selected. Differences are small but can be on the order of a whole cone and are most apparent when toggling between images (see [Visualization 3](#)).

3. Discussion

We have shown here that when time constraints are no longer a limiting factor, additional image processing steps can be applied to achieve an improved registration result, both for precise measurement of eye motion and for imaging applications. Key aspects of our approach include pre-processing contrast enhancement phases and routines for detecting highly distorted and candidate large motion frames. Perhaps one of the most different aspects of our approach compared to what has been published previously in this area is the development of a composite reference frame for the final strip level registration. This general idea was proposed originally by Stevenson (personal communication). Other methods have been proposed to correct intraframe distortions, such as the lag bias approach [17,18]. This approach computes the offset values between adjacent strips that in the absence of distortion should be close to zero and then uses this eye motion value to generate a dewarped reference frame. Though this method worked well on synthetic image sequences with simulated eye motion artifacts and on some real AOSLO image sequences, it failed when eye movements did not hold to the assumptions required for this approach, such as when one direction of drift predominates or in the case of large amplitude eye motion. Recently, a hardware based dual scanning solution has been proposed to remove distortions from scanning systems [27]. An advantage of our approach is that requires no hardware modifications and can be readily applied to existing datasets.

Our approach for generating a synthetic reference frame by averaging frames (registered at the frame level) mitigates the intraframe distortions, even in the case of anisotropic or ‘idiosyncratic’ eye movements. The principle of this approach is that because the small intraframe distortions appear at random locations across the different images, when enough of the images are registered and averaged, the information from the distorted pixels will be averaged out by the non-distorted ones dominating. It should be noted that this does not require a substantial number of frames for the averaging to mitigate the distortion. Moreover, as shown in [Visualization 4](#), our approach only requires 20–30 images (≤ 1 second at 30 fps) contributing to the synthetic reference frame to generate a near distortion-free registered and averaged image, demonstrating the potential of this algorithm for distortion mitigation in short image sequences. It should be noted that eye motion on short timescales will often be anisotropic and fail the assumptions required for the lag bias [18] approach to work effectively. An additional benefit of our synthetic reference frame approach is that it builds a reference frame with a larger field of view compared to any individual frame in the sequence, enabling our approach to also measure large amplitude eye motion.

Registration algorithms designed for eye tracking are inherently difficult to assess without a ground truth reference for comparison. This limitation forced us to utilize approaches for assessment that had limitations. For one, there were inherent differences between our method and the comparison algorithm we chose as our benchmark that do not always put them on equal footing. As we describe above, the benchmark uses a different blink detection method and there were differences in the strip parameters we used for the AOSLO data. As we show in Suppl. Fig. S4 in [Supplement 1](#), the intensity threshold-based method we used for blink detection resulted in differences between the frames labelled as blinks in comparison to the cross-correlation threshold-based method used by the benchmark [11]. It was our original intention to set all strip parameters to be identical, however, when doing so for the AOSLO data we found that this would not reflect the performance that could be achieved with the benchmark algorithm when using its default settings, so we chose to use those parameters instead. We also tested our approach using the default benchmark strip parameters and found similar performance (see Supplementary Table 1 in [Supplement 1](#)) but increased computational cost due to the larger strip size.

In terms of proportion of tracked frames or strips, we showed that our technique could achieve very high tracking rates, approaching 99% in most cases. This reflected a substantially greater proportion of eye motion that could be measured compared to the benchmark, so we were interested to know if those additional strips were tracked with the same level of precision as

those that could be successfully tracked by both techniques. Comparison of the registered image sequences between algorithms using the SD across time method showed that when frames were successfully registered in each algorithm, a similar reduction in SD across time was seen in the registered image sequence (Fig. 5). This demonstrates that the additional frames tracked by our method are registered to a similar level of accuracy at least as it is reflected by this metric. Careful inspection of the difference histograms shows that there was a small shift in the distribution towards slightly higher SD in the additional frames successfully tracked by our algorithm that were dropped by the benchmark (Fig. 5(f)). This small difference could reflect a decrease in the computation accuracy or that these frames display a higher variation in overall intensity due to large eye motion. We also observed that the proportion of frames that were untracked by the benchmark did not vary systematically with subjective image quality (see Table 1). We suspect that this is because subjective image contrast was the main visual criterion used to define the three quality levels and that image sequences with higher contrast, but greater amounts of large motion frames or blinks, could increase the number of untracked frames.

To evaluate whether our technique could successfully track both microsaccades and drifts, we segmented the microsaccades from the drift epochs in the TSLO image sequences. An example segmented motion trace is shown in Supplementary Fig. S6 in [Supplement 1](#). Both classes of fixational eye motion are also seen in the registered image sequences shown in [Visualization 1](#) and the eye trace and corresponding image sequence shown in [Visualization 2](#). Across all 30 TSLO image sequences, we found that we detected ~ 0.9 microsaccades per second, on average. This value is well within the range we expect for normal human observers based on previous studies [1,4]. This suggests that our strip-based approach successfully measured both drifts and microsaccades.

Improved strip tracking is beneficial for eye motion measurement applications, but it should be noted that keeping more frames does not always improve the quality of the final registered and averaged image for imaging applications (Fig. 6). In certain cases, averaging more strips or frames can have a detrimental effect on image quality. The need to include more data in the registered and averaged image is most important for light starved imaging applications like autofluorescence AOSLO and some forms of non-confocal AOSLO. For example, confocal AOSLO imaging of photoreceptors may only require a relatively small number of strips per pixel to be averaged to achieve a high-quality image. In those cases, one likely would only include in the final averaged images those strips that had the highest cross-correlation threshold. For those applications, we have generated an averaging and cropping tool that allows the registered image sequences averaging to be customized to the application.

For evaluating the tracking precision, we developed a tedious manual landmarking approach and deployed it on several image sequences. This was a suitable method for the TSLO data where larger vessel landmarks could be reliably marked by our human graders, but it was not useful for the AOSLO data, as we found the additional structural image detail in AOSLO made it nearly impossible for the manual graders to reliably mark the same exact structure from frame to frame. Despite this limitation, we showed that human graders could routinely detect the same structures reliably and that there were larger differences between the different algorithms than there were between the different graders.

The ability to track single cells across time within the living eye has long been an overarching goal for AO ophthalmoscopes. However, we have been limited in our ability to track cells longitudinally due to our inability to reliably reproduce the same retinal structure in the face of within-frame distortions from eye movements. Some investigators have taken the approach of warping the averaged images that they want to compare across images within an imaging session or across images taken at different timepoints. This is suitable for psychophysical testing or imaging studies on normal eyes when the retina is not expected to change between imaging sessions. However, this is unsuitable when the retinal structure is changing such as in progressing

disease or in response to treatment. The ability we have shown here to reliably reproduce the same retinal structure (Fig. 7) will facilitate all imaging applications as in all cases we seek to recapitulate the true arrangement between structures within the imaging field of view. However, it is the evaluation of cell and gene-based treatments that we think will benefit the most from this precise level of targeted cell tracking.

Despite our achievements there are several aspects that could still be improved further. For one, our blink detection approach is simplistic and fails when there are not blinks in an image sequence. Another aspect that could be improved would be to implement an automatic reference frame selection step to make this approach fully automated. We currently select the starting reference frame manually but this could easily be replaced with automatic selection either using an image quality metric [28] or an algorithm such as that described by Salmon et al. [22]. Another point worth mentioning is the precision of the NCC peak calculation, as this could be altered to improve the precision of the measurement. At present we only compute the NCC peak down to single pixel precision. We tested sub-pixel precision for the fine-scale motion tracking step and found that there was no visible difference in the resulting registered and averaged images. So, although some applications may require eye motion measurements with sub-pixel precision, we decided it was not worth the computational cost for our present applications but have left the option available in our algorithm to do so when needed. Finally, this approach does not capture the torsional eye motion that can occur during FEMs and that remains unaddressed in current techniques. In fact, we can see a rotational Glass pattern [29,30] in Fig. 2(a) induced by torsional fixational eye motion in the synthetic reference, leaving this problem to be solved in future work.

The main practical drawback to our approach is that it takes a long time in its present implementation to process the data. At present, using CUDA implementations in MATLAB only, it takes approximately 9 minutes to process 900 frames of data using the parameters outlined above. However, this long computational time can be reduced through software modifications, such as porting computationally intensive tasks to other languages or using additional hardware. Future hardware improvements predicted by Moore's Law alone should permit the present method to run at real-time rates in about 7–9 years.

4. Conclusions

We show here that our modifications to the strip-based digital image registration approach for scanned ophthalmic imaging systems accomplished our primary objectives:

- 1) Tracks the precise motion of nearly all the images in each sequence for eye motion measurement and light starved imaging applications.
- 2) Is sensitive to motion larger than the field of view of a single frame.
- 3) Reconstructs the spatial arrangement between image features consistently and accurately.

Taken together, these improvements extend the current capabilities of strip-based digital image registration for eye tracking and imaging applications. Our technique facilitates the study of fixational eye movements, an emerging area of importance for understanding early changes in diseases of the eye and brain. It will also enable the tracking of individual cells over time in health and disease to permit targeted monitoring of individual cells in response to treatment.

Funding. Department of Ophthalmology, University of Pittsburgh; BrightFocus Foundation (G2017082); National Eye Institute (R01EY030517); National Institutes of Health (CORE Grant P30 EY08098); Eye and Ear Foundation of Pittsburgh; Research to Prevent Blindness.

Acknowledgments. The authors wish to thank Scott B. Stevenson for his inspirational ideas, sharing of code, and for helpful advice. The authors also wish to thank Qiang Yang for sharing his registration software with us and for providing helpful advice.

Disclosures. The authors declare no conflicts of interest.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. M. Rucci and M. Poletti, "Control and functions of fixational eye movements," *Annu. Rev. Vis. Sci.* **1**(1), 499–518 (2015).
2. S. Martinez-Conde, J. Otero-Millan, and S. L. Macknik, "The impact of microsaccades on vision: towards a unified theory of saccadic function," *Nat. Rev. Neurosci.* **14**(2), 83–96 (2013).
3. R. M. Steinman, Z. Pizlo, T. I. Forofonova, and J. Epelboim, "One fixates accurately in order to see clearly not because one sees clearly," *Spatial Vis.* **16**(3), 225–241 (2003).
4. 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).
5. N. R. Bowers, A. E. Boehm, and A. Roorda, "The effects of fixational tremor on the retinal image," *J. Vis.* **19**(11), 8 (2019).
6. S. B. Stevenson and A. Roorda, "Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy," in *Ophthalmic Technologies XV*, Proceedings of SPIE, F. Manns, P. G. Söderberg, A. Ho, B. E. Stuck, and M. Belkin, eds. (SPIE, 2005), 5688, pp. 145–151.
7. J. B. Mulligan, in *Recovery of Motion Parameters from Distortions in Scanned Images*, J. Le Moigne, ed. (NASA Goddard Space Flight Center, 1997), pp. 281–292.
8. D. Ott and R. Eckmiller, "Ocular torsion measured by TV- and scanning laser ophthalmoscopy during horizontal pursuit in humans and monkeys," *Invest. Ophthalmol. Vis. Sci.* **30**(12), 2512–2520 (1989).
9. D. Ott and W. J. Daunicht, "Eye-Movement Measurement with the scanning laser ophthalmoscope," *Clin. Vis. Sci.* **7**(6), 551–556 (1992).
10. 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).
11. Q. Yang, J. Zhang, K. Nozato, K. Saito, D. R. Williams, A. Roorda, and E. A. Rossi, "Closed-loop optical stabilization and digital image registration in adaptive optics scanning light ophthalmoscopy," *Biomed. Opt. Express* **5**(9), 3174–3191 (2014).
12. C. K. Sheehy, Q. Yang, D. W. Arathorn, P. Tiruveedhula, J. F. de Boer, and A. Roorda, "High-speed, image-based eye tracking with a scanning laser ophthalmoscope," *Biomed. Opt. Express* **3**(10), 2611–2622 (2012).
13. 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).
14. 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).
15. J. Zhang, Q. Yang, K. Saito, K. Nozato, A. Roorda, D. R. Williams, and E. A. Rossi, "An adaptive optics imaging system designed for clinical use," *Biomed. Opt. Express* **6**(6), 2120 (2015).
16. K. E. Talcott, K. Ratnam, S. M. Sundquist, A. S. Lucero, B. J. Lujan, W. Tao, T. C. Porco, A. Roorda, and J. L. Duncan, "Longitudinal study of cone photoreceptors during retinal degeneration and in response to ciliary neurotrophic factor treatment," *Invest. Ophthalmol. Visual Sci.* **52**(5), 2219–2226 (2011).
17. P. Bedggood and A. Metha, "De-warping of images and improved eye tracking for the scanning laser ophthalmoscope," *PLOS ONE* **12**(4), e0174617 (2017).
18. M. Azimipour, R. J. Zawadzki, I. Gorczynska, J. Migacz, J. S. Werner, and R. S. Jonnal, "Intraframe motion correction for raster-scanned adaptive optics images using strip-based cross-correlation lag biases," *PLOS ONE* **13**(10), e0206052 (2018).
19. American National Standards Institute, American National Standard for Safe Use of Lasers (Z136.1-2014) (ANSI, 2014).
20. B. Leonard, M. Zhang, V. Snyder, C. Holland, E. Bensinger, C. K. Sheehy, M. Collins, A. Kontos, and E. A. Rossi, "Fixational eye movements following concussion," *Invest. Ophthalmol. Visual Sci.* **60**(9), 1035 (2019).
21. K. V. Vienola, M. Zhang, V. C. Snyder, J.-A. Sahel, K. K. Dansingani, and E. A. Rossi, "Microstructure of the retinal pigment epithelium near-infrared autofluorescence in healthy young eyes and in patients with AMD," *Sci. Rep.* **10**(1), 9561 (2020).
22. A. E. Salmon, R. F. Cooper, C. S. Langlo, A. Baghaie, A. Dubra, and J. Carroll, "An automated reference frame selection (ARFS) algorithm for cone imaging with adaptive optics scanning light ophthalmoscopy," *Transl. Vis. Sci. Technol.* **6**(2), 9 (2017).
23. A. Roorda, E. A. Rossi, Y. Zhang, S. B. Stevenson, D. W. Arathorn, C. R. Vogel, A. Parker, and Q. Yang, "Applications for eye-motion-corrected adaptive optics scanning laser ophthalmoscope videos," *Invest. Ophthalmol. Visual Sci.* **47**(13), 1808 (2006).
24. P. Mecê, E. Gofas-Salas, C. Petit, F. Cassaing, J. Sahel, M. Paques, K. Grieve, and S. Meimon, "Higher adaptive optics loop rate enhances axial resolution in nonconfocal ophthalmoscopes," *Opt. Lett.* **44**(9), 2208–2211 (2019).
25. J. M. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **327**(8476), 307–310 (1986).
26. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt. Lett.* **33**(2), 156 (2008).

27. T. Luo, R. L. Warner, K. A. Sapoznik, B. R. Walker, and S. A. Burns, "Template free eye motion correction for scanning systems," *Opt. Lett.* **46**(4), 753–756 (2021).
28. M. Mujat, R. D. Ferguson, A. H. Patel, N. Ifimia, N. Lue, and D. X. Hammer, "High resolution multimodal clinical ophthalmic imaging system," *Opt. Express* **18**(11), 11607–11621 (2010).
29. H. R. Wilson and F. Wilkinson, "Detection of global structure in glass patterns: implications for form vision," *Vision Res.* **38**(19), 2933–2947 (1998).
30. L. Glass, "Moiré effect from random dots," *Nature* **223**(5206), 578–580 (1969).