Published in final edited form as:

Opt Lett. 2021 February 15; 46(4): 753–756. doi:10.1364/OL.415285.

Template free eye motion correction for scanning systems

Ting Luo¹, Raymond L. Warner¹, Kaitlyn A Sapoznik¹, Brittany R. Walker¹, Stephen A. Burns^{1,*}

¹Optometry School, Indiana University Bloomington, 800 Atwater Ave., Bloomington, IN 47045

Abstract

Scanning imaging systems are susceptible to image warping in the presence of target motion occurring within the time required to acquire an individual image frame. In this letter we introduce the use of a dual raster scanning approach to correct for motion distortion without the need for prior knowledge of the undistorted image. In the dual scanning approach, the target is imaged simultaneously with two imaging beams from the same imaging system. The two imaging beams share a common pupil but have a spatial shift between the beams on the imaging plane. The spatial shift can be used to measure high speed events, because it measures an identical region at two different times within the time required for acquisition of a single frame. In addition, it provides accurate spatial information, since two different regions on the target are imaged simultaneously, providing an undistorted estimate of the spatial relation between regions. These spatial and temporal relations accurately measure target motion. Data from adaptive optics scanning laser ophthalmoscope (AOSLO) imaging of the human retina are used to demonstrate this technique. We apply the technique to correct the shearing of retinal images produced by eye motion. Three control subjects were measured while imaging different retinal layers and retinal locations to qualify the effectiveness of the algorithm. Since the time shift between channels is readily adjustable, this method can be tuned to match different imaging situations. The major requirement is the need to separate the two images, in our case, we used different near infrared spectral regions and dichroic filters.

Scanning imaging systems are powerful tools for real time imaging as they provide the flexibility to implement multiple forms of contrast generation, including optical coherence tomography (OCT)[1-3] and scanning laser ophthalmoscopy (SLO)[4-7] and detection modes including fluorescence[8], confocal [9-11]and dark field imaging[12, 13]. In raster scan systems, the target is typically scanned in two dimensions with a fast scan in one direction and a slow scan in the orthogonal direction. A major disadvantage of these systems is that they are susceptible to motion artifacts. That is, if the target is moving during image acquisition there is intra-frame shearing, and the resulting image is distorted. For human retinal AOSLO imaging, the fast scan is often approximately 16kHz and the slow scan is on the order of 30Hz. For OCT imaging the scanning is typically slower, but both common techniques build up an image through sequential point by point measurements.

*Corresponding author: staburns@indiana.edu.

Disclosures: The authors declare no conflicts of interest.

A standard way to correct for the intra-frame eye motion is based on the strip-wise alignment method[14, 15]. The method typically starts with choosing a template frame from a sequence of video frames and correcting the rest of the frames in the video by comparing to this chosen template. Specifically, each frame is divided into contiguous rectangular strips evenly distributed along the slow axis of the raster. These strips are compared to the template separately to determine the offset from the template at multiple positions within the raster. The measured offsets are used to compute the target motion, and motion information is then used to “correct” the frame by removing eye motion. However, in practice the “template” itself suffers frame distortion[16-20]. To improve image fidelity, hardware techniques[21] as well as more complex scan patterns[22, 23] have been used. For instance, rotating the direction of fast and slow scans for individual images, and then combining images of different scan directions based on the temporal properties of the motion can generate a more accurate OCT image. This approach is limited by the need to use resonant scanners or line scanning in higher speed imaging systems since these often use different hardware for the fast and slow scan directions and thus require increased complexity to rotate the two axes. In this letter we proposed a new algorithm based on dual scan imaging which can directly measure motion within a single frame requiring two images within the frame time. This approach, while increasing complexity somewhat, also allows the characterization of higher speed events than a typical raster scan.

We previously described the use of dual scanning for measuring events faster than the frame rate of the raster scan system, such as red blood cell motion [24, 25]. Here we show that the dual scanning technique can be used to remove the majority of inter-frame distortions of the image without requiring a template image. In the dual scanning system, light at two wavelengths enters the pupil of the imaging system at slightly different angles. This causes the imaging regions to be displaced from each other but, if the angular displacement is small, most of the regions imaged by the two channels overlap. This has two results. First, a given retinal region is imaged at two different times within each scan, first by channel 1 then channel 2. Thus, if two imaging beams are separated on the retina by approximately 50 lines in a system with a 15.1kHz rate and a pixel spacing of 1x1 μm (fig. 1), two different retinal regions separated by 50 μm are imaged simultaneously by two channels and there can be no motion shearing between these regions. Second, the region imaged by channel 2 at time 1 will be imaged again 3.3 msec later by channel 1. After estimating the displacement between the two beams, changes in the relative location of image features in these two channels represent motion which occurred during the 3.3 msec interval. Because these temporal and spatial relations extend across the entire overlapping region of the two images, motion can be computed for multiple temporal samples (for example, $s_1^{(1)}$, $s_2^{(1)}$ and $s_3^{(1)}$ are three sequential temporal samples in channel 1 fig. 1) during each scan. This allows measurement of the motion for each time pair in the image, and thus provides a way to estimate the intra-frame motions with sub-pixel accuracy. Removing these motions from the images provides an undistorted estimate of the true image within the overlap region of the two channels, without the need for comparison to an undistorted frame.

Data from the Indiana AOSLO was used to test this algorithm. The optical system has been described in detail[16, 26]. Essentially, a supercontinuum laser (NKT Photonics) is used as

the light source, being filtered at different central wavelengths for two imaging beams. A resonant scanner with an oscillation frequency of 15.1 kHz[24] generates the horizontal scan and a galvanometer generates the orthogonal vertical scan, for a video frame rate of 28 Hz. The light scattered back through the system from the retina is detected by avalanche photodiodes (APD) sampled at 30MHz and the detector signal is sampled to generate images for each channel. Two deformable mirrors are configured in a woofer tweeter configuration to correct the optical aberrations in real time[27].

For the sample data, the two imaging wavelengths were 769nm and 842nm. The pupil size of the AOSLO system is 8mm. The nominal total power is less than 350 μ W. This imaging protocol was safe for imaging time longer than 10 minutes with a safety factor greater than 10x according to the ANSI standards[28]. The focus of the longer wavelength channel was adjusted to compensate for axial chromatic aberration of the eye by moving the tip of the optical fiber axially (Thorlabs SM1Z) relative to the collimating lens.

We first compute the overall displacement ($\mathbf{r}$, $\mathbf{c}$) between the two channels to obtain the time delay for which each channel will image the same retinal location. The time interval that results in this displacement is denoted as $\mathbf{t}$. We then compute a series of temporal samples from both channels (for example, $s_1^{(1)}, s_2^{(1)}, s_3^{(1)}$ etc. in channel 1 and $s_1^{(2)}, s_2^{(2)}, s_3^{(2)}$ etc. in channel 2) spaced at a step size of $\mathbf{r}$ rows. For the current study the number of temporal samples was determined by the displacement $\mathbf{r}$, eg. 10 temporal samples for a 50 line displacement.) Pairs of temporal samples are chosen such that the i th temporal sample in channel 1 is imaged simultaneously with the i th temporal sample in channel 2. Let $S_i^{(j)}$ represent the i th temporal sample of channel j ($j = 1$ or 2). We have:

$$\Delta s_i^{(1)} = \Delta s_i^{(2)} = s_i^{(2)} - s_{i-1}^{(1)}$$

This equation is based on the requirement that the same temporal sample of both images have the same motion during the same time interval. For simplicity, in the current manuscript we calculated the displacements ($\mathbf{S}_i$) using two dimensional cross-correlation of strips centered on each temporal sample within the images to compute the displacements in each time interval. The cumulative sum of these displacements over time is used to calculate total relative displacement at each time point i from the first temporal sample, i.e., $s_i^{(2)} - s_1^{(2)}$, see fig. 1.

In fig. 1, the $s_1^{(1)}$, $s_2^{(1)}$ and $s_3^{(1)}$ are three temporal samples from channel 1, the corresponding paired temporal samples are marked on channel 2. $\mathbf{t}$ is the time interval between the two channels. The regions in the two images enclosed by the red box were acquired at the same time, similarly for pairs throughout the images and the displacement for the region in red in channel 1 and blue in channel 2 represents the motion during $\mathbf{t}$.

After estimating the displacement of each line of the frame, we applied a subpixel algorithm to move lines to correct for target motion. For missing lines, when the eye motion causes a

region to not be imaged at all, we interpolate the missing data with each missing pixel calculated as an average over (at most 4) neighboring pixels weighted by their distance.

To test the algorithm, eye movements were first simulated (Matlab, Mathwork, MA) by imposing known motions on actual images. The image used for simulation was a $4^\circ \times 4^\circ$ montage from the imaging system from which we sampled a 650×650 pixels rectangle twice with a vertical displacements. These two cropped regions are treated as the channel 1 and channel 2 images respectively. We then impose simulated motion on them by assuming 1) the target motion has a uniform acceleration, followed by a uniform deceleration during the eye movement period; 2) the lead time from the start till the end of the eye movement will be less than the frame rate of the AOSLO system (28 Hz) although this is not a requirement for the algorithm. The start and the end of the imposed motion was varied across conditions as was the acceleration.

Data were collected from the right eye of 3 participants. Subjects were instructed to look at a cross displayed via a video image display subsystem (TI, DLP 4500, Texas Instruments, Dallas TX) that was added to the system immediately in front of the eye using a pellicle. Imaging data were collected at 28 Hz in blocks of 100 consecutive video frames of 1.8×2.9 deg (520×852 pixels) imaged at the same retinal locations. The imaging location was then moved one-half of the field size and another block was acquired. This procedure was followed for between 9 and 12 blocks to montage a small rectangular retinal area. For the current study, three nominal vertical displacements were used to generate spatial offset between channels. In the absence of chromatic aberration these displacements were approximately 20, 50 and 70 lines, corresponding to 1.3 msec ($20 \text{ lines} * 15100^{-1} \text{ seconds/line}$), 3.3 msec and 4.6 msec . The actual displacements vary from the nominal displacements due to transverse chromatic aberration. To estimate the actual displacement of the beam at each retinal location, we use the median displacement computed from the 100 frame block using whole frame cross-correlation [29]. To test the algorithm, images were acquired at 2 different retinal focal depths: the photoreceptor layer and the retinal nerve fiber layer. Data were collected from 3 normal subjects (age range 27 to 35 yo). Pupils were dilated with 1 drop of 1% tropicamide prior to imaging. All research was approved by the Indiana University Institutional Review Board and adhere to the tenets of the Declaration of the Helsinki and all participants provided informed consent prior to participating in the research.

After removing the sinusoidal scanning error, we estimated the degree of similarity of whole frame images from three different operations. First, frames were rigidly translated to optimally align all video frames, that is, only interframe translation was corrected. In the second, we applied the algorithm described above to remove the impact of target motion using a single pass correction. In the third, we applied the algorithm iteratively. The multiple-pass approach was required only for eye motions that generated significant distortion even within the comparison strips. For all three operations we estimated the efficacy of the correction by comparing each video frame within an acquisition sequence to all other frames using 2D normalized cross correlation.

An example of the simulation output is shown in fig. 2. The simulation results demonstrate that the algorithm effectively corrected the small eye movements (fig. 2 (a), (b) and (c)). For

simulation of large motions (fig.2 (d), (c) and (e)) containing significant vertical motions (in the direction of the slow scan), the algorithm required multiple-pass to fully correct the motion. The single pass properly estimates and corrects most of the motion, and the multiple-pass removed the remaining motion. Also evident is a residual low pass filtering of the motion estimation. This filtering arises from both the size of the image comparison strips which average the motion across the strip, and also by the sampling of the motion at fixed time intervals which was performed at the time intervals representing the vertical displacement. The values used for simulated eye movements in fig. 2 bracket the range of most eye movements obtained in our data set with the simulated large motions being similar to what is expected in patients or individuals with poor fixation. For real data in the rest of the letter the eye movements are not as large as for the large motion simulation, both single pass and multiple-pass approaches worked equally well to increase the similarity significantly for all retinal images as shown in fig. 3.

Visualization 1 compares results for the same block of 91 video frames for control vs corrected sequences. Eye movements can be observed in the control video (left) while the corrected video (right) is more stable.

We have demonstrated that intra-frame distortions of a raster scanned image of a moving target is substantially reduced by using two imaging beams with a displacement between them. While we demonstrate this algorithm on images of the human eye, it is applicable to any scanning system and can be tuned to the expected motion by adjusting the offset between channels. This is important for the human eye because, even when fixating, the eye undergoes drift and small saccades causing distortion of the retinal images. In this paper, we used cross-correlation for determining the motion induced distortion within small retinal regions with strips oriented along the fast axis of the scanning and adjusted the height and blur size (the sigma of the gaussian filter to smooth the image) of the strips to account for difference in image features between retinal layers. Thus, image quality of single frames is important and the single frame noise contributes to the failure of the cross-correlations to reach 1.0 after correction. Additional differences between the images include the slightly different contrasts resulting from wavelength differences (fig. 1) between images. Fast torsional rotations between frames could also occur and these are not corrected within frames although because each frame acts as its own reference, rotations arising from small torsional motions would most likely appear as rotations of the whole images between frames.

Images with high contrast detail, such as confocal images were better corrected. We examined this separately by applying the algorithm to even lower contrast images, in our case, multiply scattered light images of the superficial vascular plexus, and here the algorithm failed for two reasons: 1) the lower contrast of static image details, and 2) the presence of the significant non-rigid motion which occurs because of the motion of red blood cells even within the time interval (even speeds of 1mm/sec, common in capillaries, results in cell movements of 3 microns within 3 msec.).

The ability to remove the impact of within frame target motion without reference to a template has important advantages. First, when generating montages less distortions will be

required between regions. This can be enhanced further by using the within frame eye movement measurement to select frames with minimum distortion since the motion estimation does not require an estimate of the true object. For a given retinal region, this will allow estimating the aspects of eye motion that are not captured by our algorithm, most noticeably translation and rotatory eye motions. Second, it can be readily combined with existing strip alignment methods to correct residual distortions. These uncorrected motions can occur because the motion estimate is somewhat filtered and delayed due to sampling as shown in the simulation (fig1). However, by using the frames with only small intra-frame motions as a template, a standard strip alignment algorithm can touch up the alignment, but now using an undistorted template. Using this approach, because the intraframe correction will remove most of the distortion, it should be possible to include measurements made even when there are significant motions and thus a larger proportion of the data should be useable.

There are obvious ways in which the algorithm can be improved in practice. First, the displacement of the beams on the target can be fine-tuned depending on image quality and the expected target motion. For high quality imaging but with rapid target motion, smaller beam displacements will allow more dense temporal sampling of the motion. Second, while we used a simple cross-correlation technique, more sophisticated similarity estimation techniques can be tuned to the image type. Here we used equally spaced samples in time and simple cross-correlation. If needed, feature matching algorithm could be used to identify more corresponding points and generate a more continuous estimate of motion due to the existing quality relation between all image points despite being separated by the image displacements. Third, the use of highly accurate real time stabilization of displays on the retina has proven useful for scientific purposes [30]. To date, this has been done using template images and real time computations of image displacements [18]. Because the current method does the same type of comparisons, but now using data acquired within the frame, it may be possible to implement image stabilization using the current approach but with rapid recovery following blinks and saccades.

Funding sources:

NEI R01 EY024315.

References

1. Hermann B, Fernández EJ, Unterhuber A, Sattmann H, Fercher AF, Drexler W, Prieto PM, and Artal P, "Adaptive-optics ultrahigh-resolution optical coherence tomography," *Opt Lett* 29, 2142–2144 (2004). [PubMed: 15460883]
2. Zawadzki RJ, Jones SM, Olivier SS, Zhao M, Bower BA, Izatt JA, Choi S, Laut S, and Werner JS, "Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal in vivo imaging," *Opt. Express* 13, 8532–8546 (2005). [PubMed: 19096728]
3. Zhang Y, Rha J, Jonnal R, and Miller D, "Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina," *Opt Express* 13, 4792–4811 (2005). [PubMed: 19495398]
4. Hammer DX, Ferguson RD, Bigelow CE, Iftimia NV, Ustun TE, and Burns SA, "Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging," *Opt. Express* 14, 3354–3367 (2006). [PubMed: 19516480]

5. Lu J, Gu B, Wang X, and Zhang Y, "Adaptive optics parallel near-confocal scanning ophthalmoscopy," *Opt Lett*41, 3852–3855 (2016). [PubMed: 27519106]
6. Mujat M, Ferguson RD, Iftimia N, and Hammer DX, "Compact adaptive optics line scanning ophthalmoscope," *Opt. Express*17, 10242–10258 (2009). [PubMed: 19506678]
7. R.-B. FRoorda Austin, Donnelly William J. III, Queener Hope, Hebert Thomas J., and Campbell Melanie C.W., "Adaptive optics scanning laser ophthalmoscopy," *Opt. Express*10, 8 (2002).
8. Ploem JS, "Laser scanning fluorescence microscopy," *Applied Optics*26, 6 (1987).
9. Webb RH, "Confocal optical microscopy," *Rep. Prog. Phys*59, 45 (1996).
10. Wilson T and Carlini AR, "Size of the detector in confocal imaging systems," *Opt Lett* 12, 227–229 (1987). [PubMed: 19738847]
11. Sredar N, Fagbemi OE, and Dubra A, "Sub-Airy Confocal Adaptive Optics Scanning Ophthalmoscopy," *Transl Vis Sci Technol*7, 17 (2018).
12. Elsner A, Miura M, Burns S, Beausencourt E, Kunze C, Kelley L, Walker J, Wing G, Raskauskas P, Fletcher D, Zhou Q, and Dreher A, "Multiply scattered light tomography and confocal imaging: detecting neovascularization in age-related macular degeneration," *Opt Express*7, 95–106 (2000). [PubMed: 19404374]
13. Chui TYP, Vannasdale DA, and Burns SA, "The use of forward scatter to improve retinal vascular imaging with an adaptive optics scanning laser ophthalmoscope," *Biomedical optics express*3, 2537–2549 (2012). [PubMed: 23082294]
14. Stevenson S and Roorda A, Correcting for miniature eye movements in high-resolution scanning laser ophthalmoscopy, *SPIE BiOS (SPIE, 2005)*, Vol. 5688.
15. Vogel CR, Arathorn DW, Roorda A, and Parker A, "Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy," *Opt Express*14, 487–497 (2006). [PubMed: 19503363]
16. Ferguson RD, Zhong Z, Hammer DX, Mujat M, Patel AH, Deng C, Zou W, and Burns SA, "Adaptive optics scanning laser ophthalmoscope with integrated wide-field retinal imaging and tracking," *J Opt Soc Am A Opt Image Sci Vis*27, A265–277 (2010). [PubMed: 21045887]
17. Huang G, Zhong Z, Zou W, and Burns SA, "Lucky averaging: quality improvement of adaptive optics scanning laser ophthalmoscope images," *Opt Lett*36, 3786–3788 (2011). [PubMed: 21964097]
18. Yang Q, Zhang J, Nozato K, Saito K, Williams DR, Roorda A, and Rossi EA, "Closed-loop optical stabilization and digital image registration in adaptive optics scanning light ophthalmoscopy," *Biomed Opt Express*5, 3174–3191 (2014). [PubMed: 25401030]
19. Cooper RF, Sulai YN, Dubis AM, Chui TY, Rosen RB, Michaelides M, Dubra A, and Carroll J, "Effects of Intraframe Distortion on Measures of Cone Mosaic Geometry from Adaptive Optics Scanning Light Ophthalmoscopy," *Transl Vis Sci Technol*5, 10 (2016).
20. Azimipour M, Zawadzki RJ, Gorczynska I, Migacz J, Werner JS, and Jonnal RS, "Intraframe motion correction for raster-scanned adaptive optics images using strip-based cross-correlation lag biases," *PLoS One*13, e0206052 (2018). [PubMed: 30359401]
21. Lu J, Gu B, Wang X, and Zhang Y, "High-speed adaptive optics line scan confocal retinal imaging for human eye," *PLOS ONE*12, e0169358 (2017). [PubMed: 28257458]
22. Kraus MF, Potsaid B, Mayer MA, Bock R, Baumann B, Liu JJ, Hornegger J, and Fujimoto JG, "Motion correction in optical coherence tomography volumes on a per A-scan basis using orthogonal scan patterns," *Biomed Opt Express*3, 1182–1199 (2012). [PubMed: 22741067]
23. Chen Y, Hong YJ, Makita S, and Yasuno Y, "Three-dimensional eye motion correction by Lissajous scan optical coherence tomography," *Biomed Opt Express*8, 1783–1802 (2017). [PubMed: 28663866]
24. de Castro A, Huang G, Sawides L, Luo T, and Burns SA, "Rapid high resolution imaging with a dual-channel scanning technique," *Opt Lett*41, 1881–1884 (2016). [PubMed: 27082369]
25. Warner RL, de Castro A, Sawides L, Gast T, Sapoznik K, Luo T, and Burns SA, "Full-field flicker evoked changes in parafoveal retinal blood flow," *Sci Rep*10, 16051 (2020). [PubMed: 32994535]
26. Burns SA, Tumber R, Elsner AE, Ferguson D, and Hammer DX, "Large-field-of-view, modular, stabilized, adaptive-optics-based scanning laser ophthalmoscope," *J Opt Soc Am A Opt Image Sci Vis*24, 1313–1326 (2007). [PubMed: 17429477]

27. Zou W, Qi X, and Burns SA, "Wavefront-aberration sorting and correction for a dual-deformable-mirror adaptive-optics system," *Opt Lett*33, 2602–2604 (2008). [PubMed: 19015681]
28. A. N. S. Institute, "American National Standard for the Safe Use of lasers," in ANSI Z136.1-2007 New York, (2007).
29. Winter S, Sabesan R, Tiruveedhula P, Privitera C, Unsbo P, Lundstrom L, and Roorda A, "Transverse chromatic aberration across the visual field of the human eye," *J Vis*16, 9 (2016).
30. Tuten WS, Tiruveedhula P, and Roorda A, "Adaptive optics scanning laser ophthalmoscope-based microperimetry," *Optom Vis Sci*89, 563–574 (2012). [PubMed: 22446720]

Fig. 1.

At the start of the scan (T_0) Channel 1 is imaging a vessel branch (A). Channel 2 is imaging a region above the branch. At T_1 (3msec. later) channel 1 is imaging just above the second branch (B) and channel 2 is imaging at the branch (A'). There are no distortions between B and A' as these two points are being imaged simultaneously. At time T_2 channel 2 is now imaging location B'. The difference in relative position between B and B', is a measure of the motion from the time T_1 to T_2 . This paired relation is repeated down the frame. Color coding indicates the time of the sample.

Fig. 2.

(a) & (d): Estimated shearing correction that was imposed (dashed lines) and measured by the motion correction algorithm (solid lines); (b) & (e): small/ large motion simulation of sheared input frame; (c) & (f): correction results. The inserts show a magnified view of cone photoreceptors during the simulated motion for corresponding frames with distortion of photoreceptors evident in (b) and (e) but corrected in (c) and (f).

Fig. 3.

Box plots of pairwise cross-correlation coefficient between corrected frames. For each of the three panels from left to right: “control” is the 2D normalized cross-correlation result of raw input frames; “Single pass” and “multiple-pass” algorithms show results after single and multi-pass shearing correction. Left and right groups show the cross-correlations for images which had relatively larger (top 15%) and smaller (bottom 15%) eye motions. Paired Student t-test indicates that the corrected frames have higher correlation ($p < 0.0001$) than those where only bulk translation was used.