

Retinal motion estimation and image dewarping in adaptive optics scanning laser ophthalmoscopy

Contents

1	Introduction	3
2	Mathematical Model for Scanning Data	4
3	Motion Retrieval Algorithms	4
3.1	Translational Motion Estimation Via Cross-Correlation	4
3.2	Motion Detection Using the MSC Algorithm	5
3.3	Image Preprocessing	6
4	Image Dewarping	6
5	Experimental Results	7
6	Discussion and Conclusions	9

Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy

Curtis R. Vogel

Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717-2400

vogel@math.montana.edu

<http://www.math.montana.edu/vogel/>

David W. Arathorn

Center for Computational Biology, Montana State University, Bozeman, MT 59717

dwa@cns.montana.edu

Austin Roorda

School of Optometry, University of California, Berkeley, CA 94720

aroorda@berkeley.edu

<http://vision.berkeley.edu/roordalab/>

Albert Parker

Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717-2400

parker@math.montana.edu

<http://www.math.montana.edu/parker/>

Abstract: We apply a novel computational technique known as the map-seeking circuit algorithm to estimate the motion of the retina of eye from a sequence of frames of data from a scanning laser ophthalmoscope. We also present a scheme to dewarp and co-add frames of retinal image data, given the estimated motion. The motion estimation and dewarping techniques are applied to data collected from an adaptive optics scanning laser ophthalmoscopy.

© 2005 Optical Society of America

OCIS codes: (010.1080) Adaptive Optics; (180.1790) Confocal Microscopy

References and links

1. A. Roorda, F Romero-Borja, W.J. Donnelly, T.J. Hebert, H. Queener, M.C.W. Campbell, "Adaptive Optics Scanning Laser Ophthalmoscopy," *Opt. Express* (<http://www.opticsexpress.org>) **10** (2002), pp. 405–412.
2. J. Liang, D. R. Williams, and D. Miller, "Supernormal vision and high-resolution retinal imaging through adaptive optics," *J. Opt. Soc. Am. A* **14** (1997), pp. 2884-2892.
3. R. H. Webb, G. W. Hughes, and F. C. Delori, "Confocal scanning laser ophthalmoscope," *Appl. Opt.* **26** (1987), pp. 1492-1499.
4. D.W. Arathorn, *Map-Seeking Circuits in Visual Cognition: A Computational Mechanism for Biological and Machine Vision*, Stanford University Press, 2002.
5. D. W. ARATHORN, "Computation in higher visual cortices: Map-seeking circuit theory and application to machine vision," *Proceedings of IEEE Applied Imagery Pattern Recognition Workshop* (2004), pp. 73–78.
6. D. W. ARATHORN, "From wolves hunting elk to Rubik's cubes: Are the cortices compositional/decompositional engines?" *Proceedings of AAAI Symposium on Compositional Connectionism* (2004), pp. 1–5.

7. D. W. ARATHORN, "Memory-driven visual attention: An emergent behavior of map-seeking circuits," in *Neurobiology of Attention*, Eds. L. Itti, G. Rees, and J. Tsotsos, Academic Press/Elsevier, 2005.
 8. D. W. ARATHORN, *A cortically plausible inverse problem solving method applied to recognizing static and kinematic 3-D objects*, proceedings of Neural Information Processing Systems (NIPS) Workshop, 2005
 9. D. W. Arathorn and T. Gedeon, "Convergence in map finding circuits," preprint, 2004.
 10. S.A. Harker, T. Gedeon, and C.R. Vogel, "A multilinear optimization problem associated with correspondence maximization," preprint, 2005.
 11. J.B. Mulligan, "Recovery of motion parameters from distortions in scanned images," Proceedings of the NASA Image Registration Workshop (IRW97), NASA Goddard Space Flight Center, MD, 1997.
 12. S.B. Stevenson and A. Roorda, "Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy," in *Ophthalmic Technologies XV*, edited by Fabrice Manns, Per Soderberg, Arthur Ho, Proceedings of SPIE Vol. 5688A (SPIE, Bellingham, WA, 2005), pp. 145–151.
 13. J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, 2004.
 14. <http://www.math.montana.edu/~vogel/Vision/graphics/>
 15. J. A. Martin and A. Roorda, "Direct and non-invasive assessment of parafoveal capillary leukocyte velocity," *Ophthalmology* (in press).
 16. T.N. Cornsweet and H.D. Crane, "Accurate two-dimensional eye tracker using first and fourth Purkinje images," *J. Opt. Soc. Am.* **63** (1973), pp. 921-928.
-

1. Introduction

The adaptive optics scanning laser ophthalmoscope [1] (AOSLO) is a scanning device which produces high resolution optical sections of the retina of the living eye. This instrument combines adaptive optics [2], which is a set of techniques used to measure and correct for aberrations that cause blur in retinal images, with confocal scanning laser ophthalmoscopy [3]. After adaptive optics correction, microscopic details in the human retina are directly observed. The usefulness of the AOSLO has been limited by motions of the eye that occur on time scales which are comparable to the scan rate. These motions can lead to severe distortions, or warping, of the AOSLO images. These distortions are particularly apparent in the AOSLO because, with the small field sizes that are used to achieve sufficient sampling rates for microscopic imaging (typically 1-3 degrees), the effects of the eye movements are magnified. Removing the distortions is an important step toward providing high fidelity visualizations of the retina, either as stabilized videos, or as high signal-to-noise (S:N) frames. Since the signal in a single frame is limited by safe light exposure limits, co-adding an undistorted sequence of frames is required to improve the S:N of static images. In order to correct for these distortions, one must first determine the retinal motion that has caused them.

In this paper we apply a novel computational technique known as the map-seeking circuit (MSC) algorithm [4, 5, 6, 7, 8, 9, 10] to estimate the motion of the retina from a sequence of frames of AOSLO data. For simple translational motion, one can apply standard cross-correlation techniques [11, 12]. MSC has much lower computational complexity; hence it may allow the motion to be estimated much more quickly. In addition MSC easily allows one to consider more general motions—like rotations—that may arise in AOSLO imaging. One may also be able to adapt other standard techniques from image registration [13], e.g., landmark-based registration.

Once motion estimates are available, one can compensate for the motion-induced image distortions. This is called image "dewarping". In this paper we introduce a weighting scheme that is based on linear interpolation to do the dewarping. This scheme can also be used to co-add multiple image frames in order to form image mosaics and to average out the noise.

This paper is organized as follows. In Section 2 we present a simple model for scanned image data from the AOSLO. Section 3 deals with motion estimation. We review the cross-correlation approach in the context of our data model, and we present the MSC algorithm as a computationally efficient alternative to fast Fourier transform based schemes to estimate motion based on cross-correlation. In Section 4 we present our dewarping scheme. In Section 5 we present

experimental results which demonstrate the effectiveness of both our MSC-based motion estimation scheme and our dewarping scheme. Finally, we present discussion and conclusions in Section 6.

2. Mathematical Model for Scanning Data

The first generation AOSLO device uses a pair of scan mirrors to sweep across the retina. The fast scan mirror sweeps back and forth horizontally once every 63.5 microseconds; it is attached to a vibrating piezo-electric crystal, which traces out a sinusoidal path in space-time. The scanning beam, which illuminates the retina and also drives the wavefront sensor for the AO system, is turned off for half of each scan period. Data at the extremes of the scan period are discarded and some preliminary data processing is done to remove sinusoidal effects, so all the remaining 512 processed pixels sample regions of the retina with the same horizontal extent. Each pixel subtends about .17 minutes of arc, or 0.88 microns of planar distance across the retina. The slow scan mirror sweeps vertically across the retina at a 30-hertz rate and traces out a skewed sawtooth pattern in space-time. The scan rate of the slow scan mirror is calibrated so that the pixels are nearly square. The slow scan return path takes about 10 percent of the total scan time, and the data recording during the return path are discarded.

Let $\mathbf{x} = (x, y)$ denote lateral (horizontal and vertical) position in the retinal layer being scanned (if there were no lateral displacements due to eye movements), and let $E(\mathbf{x})$ denote its reflectivity. Let $\mathbf{r}(t) = (r_H(t), r_V(t))$ represent the known raster position at time t , and let $\mathbf{X}(t) = (X(t), Y(t))$ denote the unknown lateral displacement of the retina. A continuous model for preprocessed, noise-free AOSLO scanning data is then

$$d(t) = E(\mathbf{r}(t) + \mathbf{X}(t)). \quad (1)$$

A model for recorded data is

$$d_i = E(\mathbf{r}(t_i) + \mathbf{X}(t_i)) + \eta_i, \quad (2)$$

where η_i represents noise and the t_i denote discrete pixel recording times.

Since the data are preprocessed, we can assume both horizontal and vertical scan paths, r_H and r_V , are periodic sawtooths. Thus in the absence of retinal motion, the AOSLO would measure the reflectivity of the retinal layer at discrete, equispaced points on a rectangular grid. For the first generation AOSLO, this grid is 512 pixels across by 480 pixels vertically and it is resampled each 1/30 second. With retinal motion, the sampled grid moves and is distorted so that it is no longer rectangular; see [12].

3. Motion Retrieval Algorithms

3.1. Translational Motion Estimation Via Cross-Correlation

Retinal motion in the form of a drift with constant velocity \mathbf{v} ,

$$\mathbf{X}(t) = \mathbf{X}(t_0) + (t - t_0)\mathbf{v}, \quad (3)$$

can be detected using cross-correlation techniques. If we assume the raster motion is periodic and we let τ_f denote the frame scan period, then from (3) and (1) we obtain $\mathbf{x}(t + \tau_f) = \mathbf{r}(t + \tau_f) + \mathbf{X}(t + \tau_f) = \mathbf{x}(t) + \tau_f\mathbf{v}$. Consequently, recorded image intensities in successive frames are related by

$$E(\mathbf{x}(t_i + \tau_f)) = E(\mathbf{x}(t_i) + \tau_f\mathbf{v}) + \text{noise}. \quad (4)$$

Recall that the cross-correlation between a pair of discrete rectangular images E and E' is defined to be

$$\text{corr}(E, E')_{k, \ell} = \sum_i \sum_j E(i+k, j+\ell)E'(i, j). \quad (5)$$

By finding the indices k, ℓ which maximize cross-correlation between consecutive pairs of discrete, rectangular AOSLO image frames, one can obtain an estimate for the offset $\tau_f \mathbf{v}$ in (4), and then extract the drift velocity \mathbf{v} .

The above cross-correlation approach may be used to estimate translational motion between non sequential image frames; pixels that are recorded at integer multiples of the frame scan rate correspond as they do in the sequential frame case, and offsets are integer multiples of $\tau_f \mathbf{v}$.

There are several shortcomings to the basic cross-correlation approach outlined above. The motion within each frame period is typically not even close to a constant drift. This problem is addressed in Section 3.3 below. Another shortcoming is due to the fact that the reflectivity E at a particular point may vary with time. This problem may be dealt with by changing the reference frame (the fixed frame against which the other frames are cross-correlated) when large changes in reflectivity are detected.

The standard approach to cross-correlation of discrete images relies on the fast Fourier transform (FFT). Stevenson and Roorda have made use of FFT-based correlation for retinal motion detection [12].

3.2. Motion Detection Using the MSC Algorithm

MSC is a method for discovering the “best” transformation T , from a given class of transformations, that maps a match image to a reference image. In the context of maximizing the cross-correlation (5), the transformations are translations and can be discretely parameterized as

$$T_{k,\ell} E(i, j) = E(i + k, j + \ell).$$

Each of the translations can be decomposed into a horizontal shift by k pixels, which we denote by $T_k^{(1)}$, followed by a vertical shift of ℓ pixels, denoted by $T_\ell^{(2)}$, so that

$$T_{k,\ell} = T_\ell^{(2)} T_k^{(1)}. \quad (6)$$

We introduce the notation

$$\langle\langle E, E' \rangle\rangle = \sum_i \sum_j E(i, j) E'(i, j), \quad (7)$$

so that the cross-correlation (5) can be expressed as

$$\text{corr}(k, \ell) = \langle\langle T_\ell^{(2)} T_k^{(1)} E, E' \rangle\rangle. \quad (8)$$

A key idea underpinning MSC is that of superposition. Rather than explicitly selecting discrete indices (k, ℓ) to maximize the cross-correlation, we select coefficient, or “gain”, vectors $\mathbf{g}^{(1)} = (g_1^{(1)}, g_2^{(1)}, \dots)$ and $\mathbf{g}^{(2)} = (g_1^{(2)}, g_2^{(2)}, \dots)$ which maximize the extended cross-correlation,

$$\text{corr}(\mathbf{g}^{(1)}, \mathbf{g}^{(2)}) = \langle\langle \left(\sum_\ell g_\ell^{(2)} T_\ell^{(2)} \right) \left(\sum_k g_k^{(1)} T_k^{(1)} \right) E, E' \rangle\rangle. \quad (9)$$

Note that when one picks the gain vectors to be standard unit vectors (all zeros except a single “1” entry), the extended cross-correlation reduces to the standard cross-correlation.

The MSC algorithm is an iterative scheme which makes use of efficiently computed gradients of the extended cross-correlation (9) to very quickly find the maximizer of the standard cross-correlation (8). Whereas standard approaches to maximizing correlation make use of the fast Fourier transform and have computation expense proportional to $N \log N$, where N is the number of pixels, the computational expense of MSC is proportional to N , and the proportionality constant is quite small. MSC’s cost can be dramatically reduced if one can obtain a “sparse

encoding” of the information in the images. By this we mean a very compact representation which requires minimal storage.

MSC can easily handle more general classes of transformations. For instance, if one wishes to consider rotations as well as translations, one can simply add a rotational “layer” by adding another term $T_m^{(3)}$ to the decomposition (6). The $T_m^{(3)}$ represents a discretization of the rotations to be considered.

3.3. Image Preprocessing

As we noted at the end of Section 3.1, retinal motion within the frame scan period is much more complex than a constant drift. However, during short time intervals, the motion can be well-approximated by drift. Since the horizontal scan rate is several orders of magnitude faster than the vertical scan rate, we first partition each image frame into several patches that extend horizontally across the frame but are some fraction (typically 1/8 or 1/16) of the vertical extent of the frame. Hence the patches are typically 512 pixels across by 60 or 30 pixels vertically. We then apply correlation between corresponding patches in the match and the reference frames to obtain approximations to the motion within frames.

If a Fourier-based cross-correlation scheme is used, one must zero pad the patches to minimize motion artifacts due to periodic wrap-around. With MSC, we take the match patch to be smaller than the reference patch. The sum in eqn (7) is then taken only over pixels in the smaller patch. With large retinal motions, there is a problem when the reference and match patches have relatively small overlap. Since the patches are smaller in the vertical direction, we shift the boundaries of match patch up or down. The amount of this up or downward shift is determined by a “look-ahead” strategy based on motion predicted from previous frames. Vertical motions larger than the patch height result in the loss of top and bottom patches. With very larger motions we find it necessary to change the reference image.

Some additional image preprocessing is required to increase the MSC convergence rate and to decrease the chance that MSC might converge to a spurious local maximizer. With AOSLO data, we decompose each image frame into several binary image “channels”, consisting for example of pixels of high intensity in one channel, middle intensity pixels in a second channel, and low intensity pixels in a third channel.

By recording only the locations of the nonzero pixels in the binary images, we can dramatically reduce the storage requirements of the MSC algorithm. Moreover, operation counts can be dramatically reduced by computing on only the nonzero pixels.

4. Image Dewarping

Let E denote the reference image and let E' denote the “match” image onto which the reference image is mapped via a transformation T . We map the match image back to the reference image using the inverse transformation to obtain a new image,

$$E''(\mathbf{x}) = E'(\mathbf{x}'), \quad \text{where } \mathbf{x} = T^{-1}\mathbf{x}'.$$

Problems arise when the image data are discrete. If a point \mathbf{x} lies on a regular reference grid, then $\mathbf{x}' = T\mathbf{x}$ is unlikely to lie on this same regular grid. In a reciprocal manner, the match image E' is implicitly collected on a rectangular array. Points \mathbf{x}' on this “match array” will not map back to the reference grid. See Fig. 1 for an illustration.

In order to obtain intensity values at points on the reference grid, we apply the following weighting scheme, which is based on bilinear interpolation. We first initialize (i.e., set components to zero) a pair of arrays— $Earray$, which will contain intensity values each point in the reference grid, and $Warray$, which contains weights at each point in the reference grid. We then

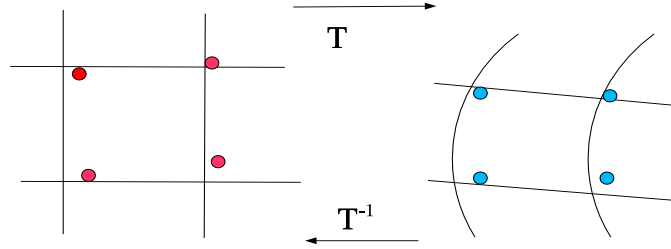


Fig. 1. Illustration of transformation effects. Under the transformation T , straight lines in the rectangular grid on the left map to curved lines on the right. Under the inverse transformation T^{-1} , equispaced grid points on the right (blue dots) map back to non equispaced points on the left (red dots).

sweep through the points in the match grid. For each point \mathbf{x}'_i in the match grid, we apply the inverse transformation to obtain $\mathbf{x}''_i = T^{-1}\mathbf{x}'_i$. We next determine into which rectangle in the reference grid \mathbf{x}''_i falls. For each of the four reference grid points \mathbf{x}_j at the corner of this rectangle, we do the following: If $Warray(j) = 0$, we reset intensity value $Earray(j) = E^l(\mathbf{x}'_i)$, and we reset the weight $Warray(j) = \frac{|x''_i - x_j|}{h_x} \frac{|y''_i - y_j|}{h_y}$. Here x_j and y_j denote the x- and y-components of \mathbf{x}_j , and similarly for x''_i and y''_i , and h_x and h_y denote the mesh spacings in the x- and y-directions. On the other hand, if the current weight $Warray(j) \neq 0$, we compute a new weight as before, but we reset the intensity value $Earray(j)$ to be a convex combination of the old intensity value and $E^l(\mathbf{x}'_i)$, where the coefficients in the convex combination are determined by the current and new weights. We then add the new weight to the current weight $Warray(j)$.

5. Experimental Results

The video clip in Fig. 2 shows a 24 frame sequence (.8 seconds) of AOSLO data in which the sinusoidal warp has been removed (see beginning of Section 2). Note that the frames have been clipped from the original 512×480 pixel size to the smaller 350×350 pixel size. In this video one observes that the predominant motion is a somewhat erratic downward drift followed by a jerky, short-duration, large-amplitude “snap-back” (this is called a micro-saccade), followed by more downward drift. One can also observe a high temporal frequency, low amplitude jitter, known as tremor, superimposed onto the larger motions.

To detect motion relative to a fixed reference frame (taken to be the first frame), we implemented the patch-based cross-correlation scheme using the MSC algorithm, which was described in Section 3. Fig. 3 shows the horizontal and vertical motion estimates that were obtained. The positive direction in subplot (a) corresponds to motion to the right in the video, while the positive direction in subplot (b) corresponds to downward motion in the video. One

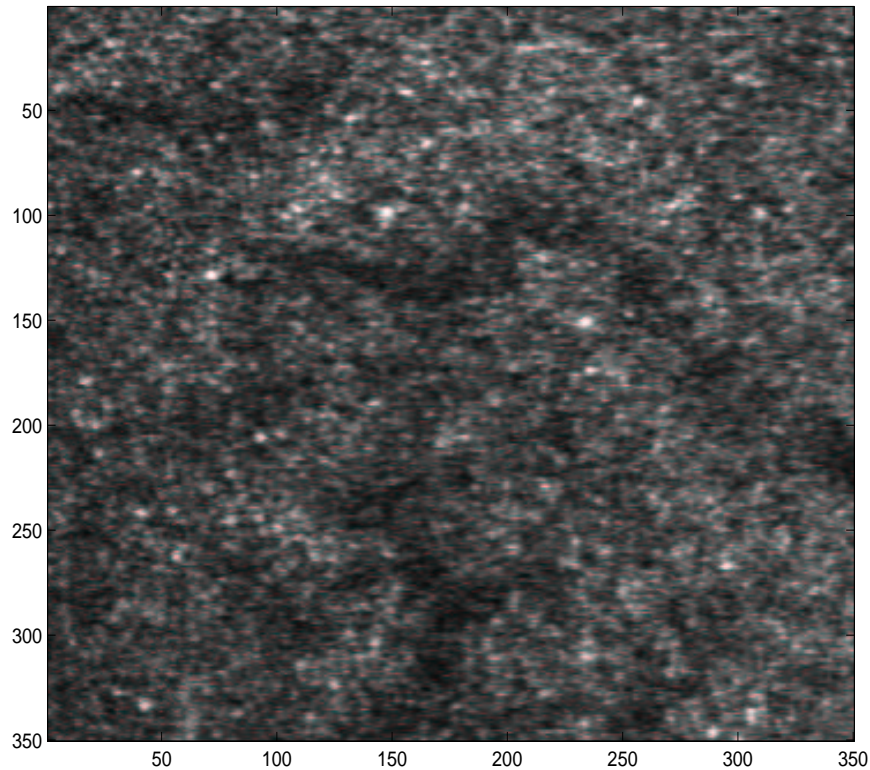


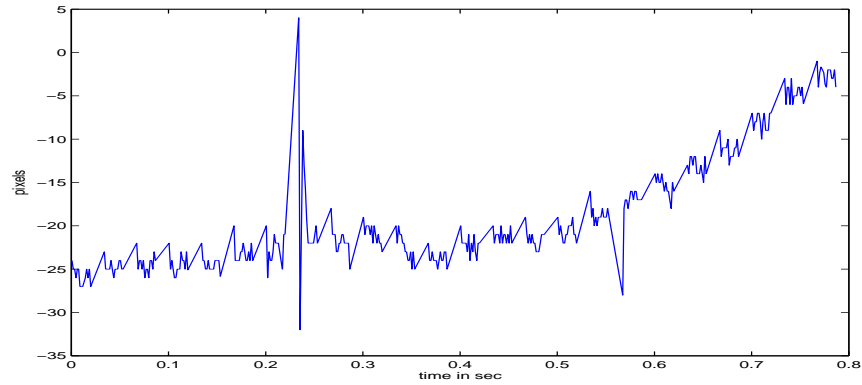
Fig. 2. Sample frame from the raw video clip SLD-AR.avi. This clip consists of 24 image frames and the file size is 3.1 MB. The image size is 350×350 pixels, or 1.02×1.02 degrees, or 300×300 microns. The fovea is located 400 microns up and to the left of the frame.

can clearly see the downward drift and the microsaccade in subplot (b). In both the video and subplot (a), one can also see a drift to the right near the end of the recording period. Note that we have used simple linear interpolation to fill in the missing motion due to lack of recorded data during the return time, or “flyback”, of the slow scan mirror and due to the loss of data from clipping.

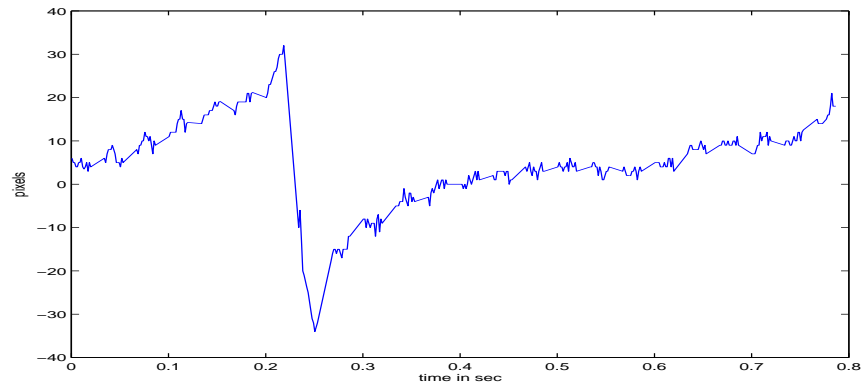
Given the motion estimates from Fig. 3, we apply the image dewarping scheme described in Section 4. Results obtained from this approach are shown in the video clip in Fig. 4. Note that individual photoreceptors, which appear as bright spots, are nearly stationary in the dewarped video. The motion of the retina is now manifested in the movement of the boundary.

We have set up a web link [14] on which we have posted a pair of much longer video clips, `OC_foveal_PRs.avi` and `OC_dewarp.avi`, showing a sequence of 300 frames (10 seconds) of raw AOSLO image data (with only sinusoidal warp removed) and the corresponding image frames after motion estimation and dewarping. Again the individual photoreceptors in the dewarped video clip are nearly stationary.

In both the raw and the dewarped videos, one can see bright “blobs”, which are leukocytes (white blood cells), moving through dark regions which are capillaries. The invention of the AOSLO has made it possible to noninvasively measure blood flow in the retina [15], albeit manually. The image stabilization in the dewarping videos may facilitate automated blood flow measurements.



(a) Horizontal Motion



(b) Vertical Motion

Fig. 3. Horizontal and vertical motion estimates obtained from AOSLO data. One pixel corresponds to .17 minutes or arc, or .88 microns of planar distance across the retina. The .8 second duration of the motion corresponds to 24 frames of AOSLO data.

Fig. 5 shows the effect of co-adding dewarped image frames. The top image in this figure is of a single frame of AOSLO data from the video clip in Fig. 2. The bottom image in Fig. 5 was obtained by co-adding the dewarped frames that appear in the second video clip (Fig. 4). An examination of the region 100 pixels down and 100 pixels across from the upper left corner of co-added image reveals a honeycomb structure known as the cone mosaic. This feature cannot be seen in the single raw image frame.

6. Discussion and Conclusions

By applying the MSC-based motion estimation scheme presented in Section 3 to AOSLO data, we were able to estimate retinal motion on time scales of $1/(30 \times 16) \text{ sec} \approx 2 \text{ millisecond}$. Given the motion estimates, we were then able to apply the dewarping scheme presented in Section 4 to remove motion-induced artifacts. We were also able to co-add dewarped data frames to

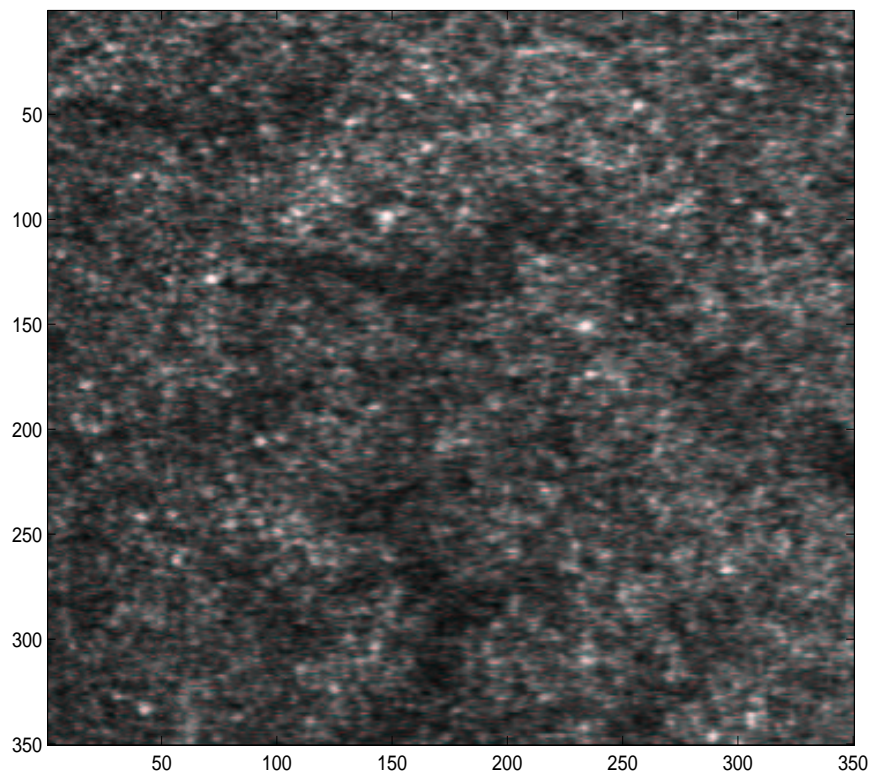


Fig. 4. Sample frame from dewarped video clip SLD-AR-dewarp.avi. This clip consists of 24 image frames and the file size is 3.5 MB. The image statistics are the same as in Fig. 2.

enhance image quality to the point where previously unobservable features like the cone mosaic could be seen.

The fact that there was very little residual motion in the dewarped video suggests that the motion estimates are quite accurate. Additional evidence of the accuracy of the motion estimation scheme comes from a study conducted by Dr. Scott Stevenson at the University of Houston School of Optometry. Stevenson obtained retinal motion estimates using (i) our MSC-based cross-correlation approach applied to AOSLO data; (ii) a standard Fourier-based cross-correlation approach applied to the same AOSLO data; and (iii) simultaneous recordings of eye motion taken from a dual-Purkinje eye tracker [16]. Stevenson found the MSC and the Fourier-based motion estimates to be in very close agreement. These in turn agreed fairly well with the Purkinje data. Discrepancies with the Purkinje measurements can be accounted for by the fact that the Purkinje measures the motion of the lens of the eye rather than the motion of the retina itself.

Note that schemes that are based on correlating patches within successive frames can detect only *relative motion*. For instance, one can “freeze” the reference frame (which was the first frame in the case presented above) and compute motion relative to the reference frame. However, the retina moves during the 1/30 second scan period during which the reference frame is recorded. Consequently, the estimated motion will be in error, or “biased”, by an amount that depends on the retinal motion during the scan of the reference frame.

To obtain the results presented in Section 5 we assumed that the average of the true motion,

$$\frac{1}{T} \int_0^T \mathbf{X}_{\text{true}}(t) dt, \quad (10)$$

tends to a constant for large T . Consequently, if a nonconstant reference frame bias is present in the estimated motion $\mathbf{X}(t)$, one can then extract the bias from the fact that

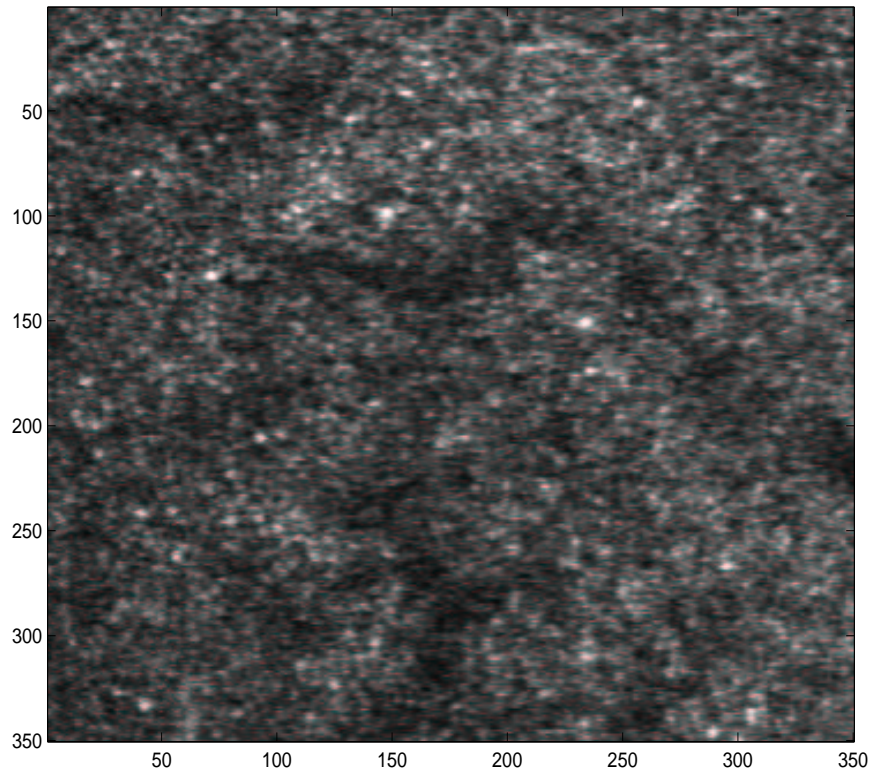
$$\frac{1}{N} \sum_{n=1}^N \mathbf{X}(t + n\tau_s) \approx \mathbf{X}_{\text{bias}}(t)$$

for large N . Here τ_s is the frame scan period, and $\mathbf{X}_{\text{bias}}(t)$, $0 \leq t \leq \tau_s$, is the (nonconstant) reference frame bias. The corrected motion for the n th scan period is then $\mathbf{X}(t + n\tau_s) - \mathbf{X}_{\text{bias}}(t)$, $0 \leq t \leq \tau_s$.

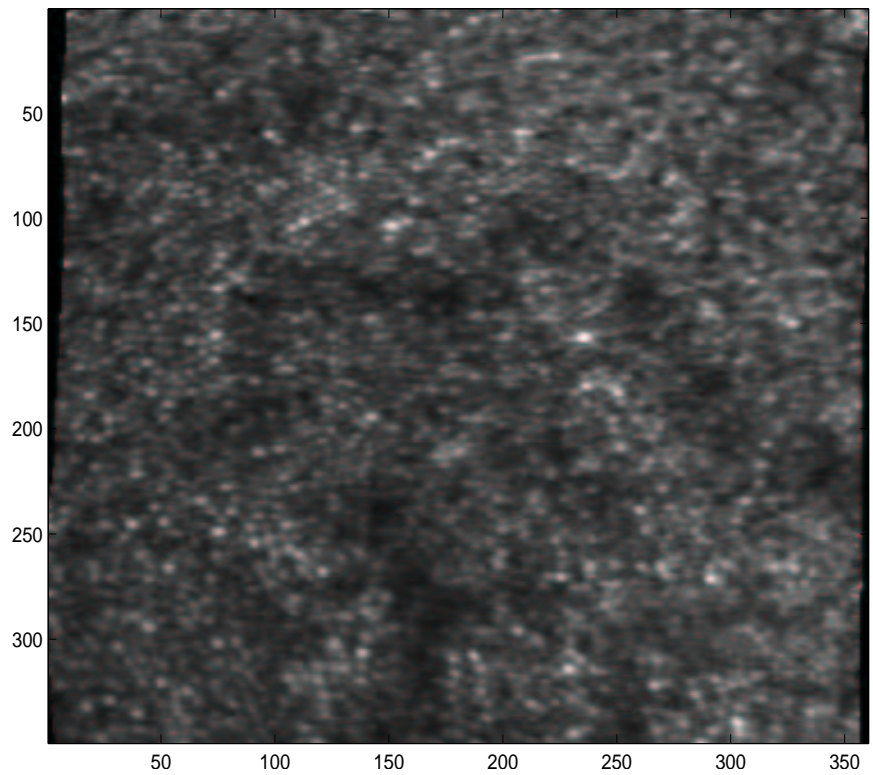
Acknowledgments

This research was supported in part by the Center for Adaptive Optics, an NSF-supported Science and Technology Center, through grant AST-9876783. Additional support comes from the NIH Bioengineering Partnership at the University of Rochester Center for Visual Science through grant EY014365.

We wish to thank Dr. Scott Stevenson of the College of Optometry at the University of Houston for supplying us with his simultaneously recorded AOSLO data and Purkinje tracker data, which we used to validate our motion estimation algorithm.



(a) Raw Image



(b) Co-added Image

Fig. 5. Raw image (top) and co-added image (bottom) obtained from AOSLO data. Image statistics are the same as in Fig. 2. Note the honeycomb structure known as a cone mosaic in the co-added image.