Retinal Motion Tracking in Adaptive Optics Scanning Laser Ophthalmoscopy

Curtis R. Vogel[1], **David Arathorn**[1], **Albert Parker**[1], **and Austin Roorda**[2]

[1] *Montana State University, Bozeman, MT 59717-2400 and* [2] *University of California, Berkeley, CA 94720*

vogel@math.montana.edu

Abstract: We apply a novel computational technique known as the map-seeking circuit algorithm to track the motion of the retina of eye from a sequence of frames of data from a scanning laser opthalmoscope.

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

1. Introduction

The adaptive optics scanning laser opthalmoscope [1] (AOSLO) is a scanning device which produces high resolution optical sections of the retina of the living eye. This instrument combines adaptive optics, to correct for aberrations of the eye, with confocal microscopy. Unfortunately, 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. In order to correct these distortions, one must first determine the retinal motion.

In this paper we apply a novel computational technique known as the map-seeking circuit (MSC) algorithm [2] to track the motion of the retina from a sequence of frames of AOSLO data. Simple translational motion can be detected using standard cross-correlation techniques [3, 4]. MSC allows one to consider more general motions, like rotation, compression, and shear, that arise in AOSLO imaging.

2. Mathematical Model for Scanning Data

Assume the retina is locally planar, let $\mathbf{x} = (x, y)$ denote position, and let $E(\mathbf{x})$ denote reflectivity. Let $\mathbf{r}(t) = (r_H(t), r_V(t))$ represent raster position at time t, and let $\mathbf{X}(t) = (X(t), Y(t))$ denote displacement of the object (the retina). A continuous model for noise-free data is then

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

A model for recorded data is

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

where η_i represents noise and the pixel recording times $t_i = i\Delta t$ are equispaced. The raster path has a fast horizontal component r_H which is periodic with period $n_H \Delta t$, where n_H is on the order of 10^3. It also has a slower vertical component r_V, which has period $n_V \tau = n_V n_H \Delta t$, where n_V is also on the order of 10^3. Hence the total time required to capture an "image frame" is on the order of $10^6 \Delta t$.

Fig. 1 shows an idealized model for the raster path for the AOSLO. The image data d_i can be ordered into $n_H \times n_V$ image arrays, or frames. Fig. 2 shows a pair of consecutive AOSLO image frames. The AOSLO pixel recording time Δt is .06 microsec; the frame recording rate is 30 frames per second; and each pixel subtends about .17 minutes of arc, or one micron of planar distance across the retina.

Since raster motion is periodic with period $T = n_V n_H \Delta t$, differences between successive image frames should be due primarily to varying object displacement $\mathbf{X}(t)$. (We assume object illumination varies much more slowly than frame recording time). Our goal is then to use a sequence of frames to extract the change in displacement relative to the position of a selected reference frame.

3. The Motion Retrieval Algorithm

We model OSLO image frames E, E' as vectors in inner product spaces $\mathcal{H}_1,\ \mathcal{H}_2$, respectively. Given a transformation T mapping $\mathcal{H}_1$ into $\mathcal{H}_2$, we define the *correspondence* between E and E' associated with T to be the inner product

$$\langle T(E), E' \rangle_{\mathcal{H}'}. \tag{3}$$

Our goal is to find the transformation which maximizes the correspondence over a set of linear transformations which can be decomposed as

$$T = T^{(L)}_{i_L} \circ \cdots \circ T^{(1)}_{i_1}, \tag{4}$$

where for each "layer" ℓ between 1 and L we have $i_\ell \in \{1, 2, \ldots, n_\ell\}$. We can then quantify the correspondence in terms of a multi-dimensional array with components

$$c(i_1, \ldots, i_L) = \langle T^{(L)}_{i_L} \circ \cdots \circ T^{(1)}_{i_1}(E), E' \rangle_{\mathcal{H}'}, \tag{5}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}'}$ denotes inner product in $\mathcal{H}'$. Thus correspondence maximization is equivalent to computing indices $(i_1, \ldots, i_L)$ which maximize the $n_1 \times n_2 \times \cdots \times n_L$ array in Eqn. (5).

When E and E' are rectangular arrays of pixel intensities and the allowed transformations are cyclic horizontal and vertical shifts, the correspondence array in Eqn. (5) is simply the cross-correlation between E and E'. This framework allows us to consider other transformations, like rotations, in addition to horizontal and vertical shifts in the decomposition (4).

The MSC algorithm [2] can be viewed as a neurobiologically motivated scheme for computing a transformation which maximizes the correspondence. A key idea underlying MSC is that of superposition. At each layer ℓ one forms a superposition, or linear combination, of all the transformations $T^{(\ell)}_i$ available at that layer. MSC then iteratively "culls" the superpositions, i.e., it eliminates all but one coefficient in the linear combination at each layer. Due to lack of space we do not provide additional algorithmic details here, but instead refer the reader to [2, 5, 6].

A brute-force direct search of the correspondence array would first require the application of $n_1 n_2 \ldots n_L$ distinct component transformations $T^{(\ell)}_i$ to assemble the array in (5). MSC is an iterative algorithm that requires only $2(n_1 + n_2 + \ldots + n_L)$ transformations per iteration. Hence, MSC's cost per iteration may be orders of magnitude less than the cost of correspondence maximization via a direct search.

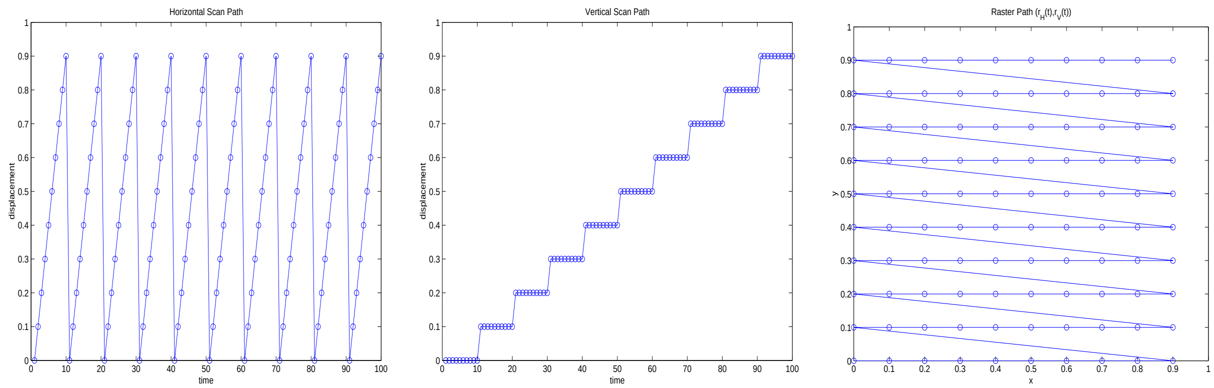
A certain amount of image preprocessing is required to increase the MSC convergence rate and decrease the chance that MSC might converge to a spurious local maximizer. With AOSLO data, we decompose each image frame into several binary image "channels", consisting for example of pixels of high intensity in one channel, middle intensity pixels in a second channel, and low intensity pixels in a third channel. In addition, rather than looking for transformations which map frame to frame, we first break the frames into patches and we then map a patch in one frame to an associated patch in the second frame. Since the horizontal scan rate is several orders of magnitude faster than the vertical scan rate, we take patches which are long in the horizontal direction and short in the vertical direction.

4. Experimental Results

Fig. 3 shows horizontal and vertical motion tracks that were obtained by applying the MSC algorithm to AOSLO scanning data. We experimented with both the patch width (size of the patches in the vertical, or slow scan direction) and with the number of layers. Two MSC layers were required to characterize the motion in terms of vertical and horizontal translations that varied from patch to patch. Four layers were required to characterize motion in terms of translation, compression, and shear. These graphs indicate that the estimated motion is consistent as the number of patches varies. With four MSC layers we were somewhat better able to track retinal motion through rapid, relatively high amplitude motions known as micro-saccades. We have been able to get good agreement between our motion estimates and motion tracks recorded with a separate device called a dual Purkinje image motion tracker [7].

References

1. A. Roorda, F Romero-Borja, W.J. Donnelly, T.J. Hebert, H. Queener, M.C.W. Campbell, "Adaptive Optics Scanning Laser Ophthalmoscopy", Optics Express, **10** (2002), pp. 405–412.
2. D.W. Arathorn, *Map-Seeking Circuits in Visual Cognition: A Computational Mechanism for Biological and Machine Vision*, Stanford University Press, 2002.
3. 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.
4. S.B. Stevenson and A. Roorda, "Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy", Proc. SPIE 5688, Ophthalmic Technologies XV, 2005.
5. D. W. Arathorn and T. Gedeon, "Convergence in map finding circuits", SIAM J. Dynamical Systems, submitted, 2004.
6. S.A. Harker, T. Gedeon, and C.R. Vogel, "A multilinear optimization problem associated with correspondence maximization", preprint, 2005.
7. 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.



(a) (b) (c)

Fig. 1. Idealized AOSLO raster path. Subplot (a) shows horizontal component; (b) shows vertical component; (c) shows composite raster path.

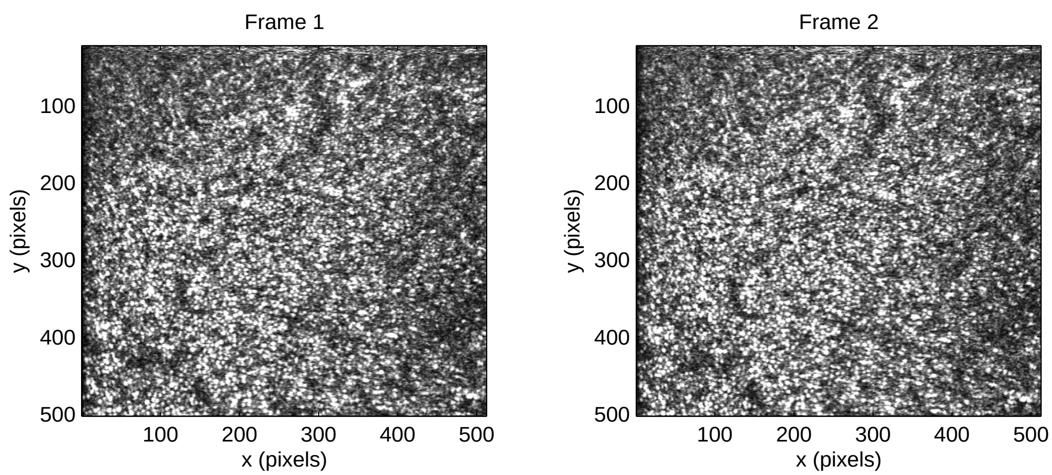
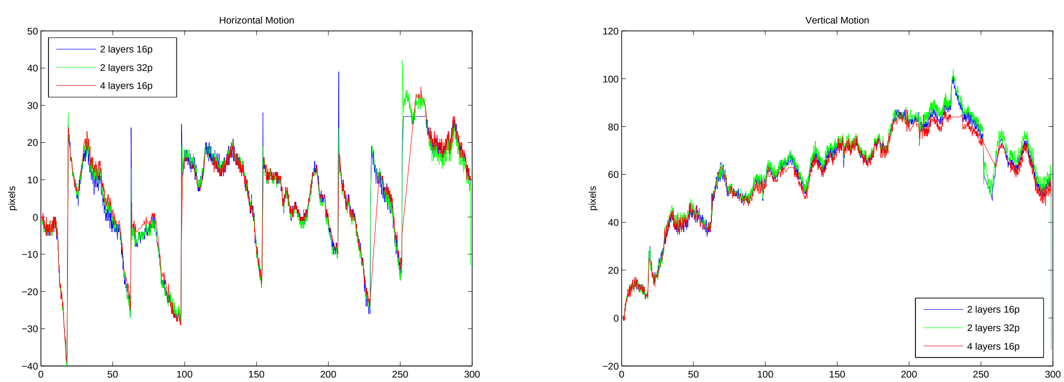


Fig. 2. Consecutive frames of AOSLO data.



(a) (b)

Fig. 3. Horizontal and vertical motion tracks obtained from AOSLO data.