

Registration of 2D Images from Fast Scanning Ophthalmic Instruments

Alfredo Dubra and Zachary Harvey

Flaum Eye Institute, University of Rochester
Rochester, NY 14642-0314, USA
`adubra@cvs.rochester.edu`

Abstract. Images from high-resolution scanning ophthalmic instruments are significantly distorted due to eye movement. Accurate image registration is required to successfully image subjects who are unable to fixate due to retinal conditions. Moreover, all scanning ophthalmic imaging modalities using adaptive optics will benefit from image registration, even in subjects with good fixation and anaesthetized animals. Transformation functions used to map two images could in principle be very complex. Here, we show that when the scanning in ophthalmic instruments is sufficiently fast with respect to the speed of involuntary eye movement, these mapping functions become the addition of a linear term and a single variable function. Then, based on experimental data on eye movement amplitude and speed of the fixating eye, minimum sampling frequencies for these instruments are discussed. Finally, a simple method for estimating the image transformation functions by taking advantage of the finite bandwidth of the motion signals is presented.

1 Introduction

The development of ophthalmic adaptive optics (AO) in recent years has led to a new generation of high-resolution retinal imaging instruments [1–11]. The unprecedented resolution of these instruments allows for *in vivo* non-invasive imaging of retinal cell mosaics [1, 9, 10, 12]. Within this family of instruments, scanning devices such as the AO optical coherence tomograph (AO-OCT) [3–8] and the AO scanning laser ophthalmoscope (AOSLO) [7, 9–11] produce images with better signal-to-noise ratio (SNR), lateral resolution and axial sectioning than flood-illuminated cameras [1, 2]. On the other hand, scanning instruments are susceptible to image distortion due to eye movement. The distortion scales with the magnification of the instrument, which is an order of magnitude greater in AO ophthalmic instruments than in commercial clinical devices.

When using high magnification to view the retina *in vivo*, light safety standards [13] severely restrict the number of photons that can be delivered within a certain period of time. For example, when visualizing lipofuscin in retinal pigment epithelial cells using single-photon fluorescence [10, 12] at safe light levels, over 99% of the recorded image pixels had no signal. The resulting images, dominated by photon noise, do not contain enough information for registration. This

led to the development of a dual-imaging method, in which a second imaging channel simultaneously recorded a reflectance signal with much higher SNR. The motion in the image was then estimated as a rigid translation using the normalized cross-correlation (NCC) in the reflectance image series, and compensated for in the matching fluorescence image series. This method produced fluorescence images in which sub-cellular structures could be identified after averaging > 1000 frames. The image registration using rigid translations was successful with anaesthetized animals and some human subjects. However, this method does not produce acceptable images in subjects that are unable to fixate. More importantly, the light safety limits are based on data from healthy subjects, and very little is known about how the damage thresholds change with eye disease, in particular, from the photochemical point of view. Therefore, in order to reduce the subjects' exposure to light and increase SNR, more advanced registration methods such as the one discussed here are required.

The next section presents the requirements for adequate sampling in scanning ophthalmic instruments. This is followed by the introduction of a mathematical model used for fitting the retinal motion transformation functions based on physical arguments. Section 4 details how the parameters of the model can be estimated from the image pairs, and illustrates the improvement that can be obtained from this registration with real data.

2 Eye Movements and Sampling Frequency

When subjects are being imaged in a high-magnification scanning ophthalmic instruments they suppress all voluntary eye movement by fixating onto a target. Fixation, however, does not suppress involuntary eye movements like tremor, drift and micro-saccades [15–19]. These eye movements have amplitudes comparable to or greater than the Rayleigh criterion for lateral resolution [20]. In a human eye with an 8 mm pupil diameter and a wavelength (λ) in the 0.5–1.0 μm range, the resolution limit is 16–31 arcsec. Tremor is a fast jerky random movement with amplitudes of 10–20 arcsec at frequencies of up to 200 Hz [15–17, 21]. According to the Whittaker-Nyquist-Shannon theorem, in order to reproduce tremor faithfully, the sampling rate should be at least 400 Hz. Drift is a continuous motion in a random direction, with speeds that could vary from 5–20 arcmin/s in normal subjects [18, 21], to 200 arcmin/s in subjects with diseased

Table 1. Temporal frequencies required to adequately sample retinal motion in scanning ophthalmic instruments due to involuntary movement in the fixating human eye

Eye movement	Minimum sampling frequency (Hz)
Drift (normal subjects, $\lambda = 0.5\mu\text{m}$)	150
Tremor	400
Drift (diseased retina, $\lambda = 0.5\mu\text{m}$)	1500
Micro-saccade	5000

retinas [19]. According to the resolution criterion for $\lambda = 0.5\mu\text{m}$ adequate sampling requires a minimum frequency of 150 Hz for normal subjects and 1.5 kHz for subjects with poor fixation. Microsaccades are short fast eye movements that compensate for the gaze drift, with amplitudes between 1 and 20 arcmin [21] and speeds in the order of 10 deg/s. Adequately sampling this type of eye movement would require sampling frequencies greater than 5 kHz. Note that the sampling frequencies mentioned above, and summarized in table (1), assume the absence of noise in the sampling process.

3 Transformation Function Model

The registration of 2D images requires transformation functions that map the coordinates (x, y) in a given (current) frame onto the corresponding coordinates (X, Y) in another (reference) frame. These mapping functions should be part of a mathematical model that adequately describes the imaging process,

$$\begin{aligned} X &= f_x(x, y; p_1, \dots, p_n), \\ Y &= f_y(x, y; q_1, \dots, q_n), \end{aligned} \quad (1)$$

with the model parameters p_j and q_j estimated using a number of control points (x_i, y_i) , (X_i, Y_i) with $i = 1, \dots, N$.

In scanning ophthalmic instruments, the image distortion due to eye movement is minimized by scanning the retina as fast as both the technology and light safety limits allow [13]. Fast scanners (y -axis) currently used for creating 2D rectangular images are either rotating polygon mirrors or resonant galvanometric optical scanners. On the other hand, the slow scanner (x -axis) is usually a non-resonant galvanometric optical scanner. The image line capture rate along the x -axis is currently in the order of 8–16 kHz. For resonant scanners the line capture occurs in the semi-cycle of the mirror oscillation, usually referred to as the forward scanning. Thus, the line acquisition time is around 20–40 μs . According to the eye movement data from the previous section, retinal features would displace about 0.01 μm due to tremor, 0.004 – 0.04 μm due to drift and 0.12 μm due to microsaccades during the recording of a single image line. Given that these values are an order of magnitude smaller than the resolution limit, it is reasonable to assume that the retinal motion can be neglected within each image line. If, in addition, it is assumed that involuntary eye rotations can be neglected, then the form of the transformation functions simplifies to

$$X = x + \epsilon_x(x), \quad (2)$$

$$Y = y + \epsilon_y(x). \quad (3)$$

Note that the linear terms have unit amplitude because the scanning speed does not change across frames. The functions ϵ_x and ϵ_y represent eye motion along the corresponding axes, and describe the compression and shear observed in the retinal images, respectively.

The eye is a mechanical system with inertia and therefore, the functions ϵ_x and ϵ_y can be modeled as having finite bandwidths, with an associated maximum frequency f_{\max} . If the sampling frequency achieved with every image line in a scanning system is greater than $2f_{\max}$, then ϵ_x and ϵ_y can be described as a sum of cosines with increasing frequencies. The amplitude of those cosines can be calculated using the discrete cosine transform II (DCT), retaining only the coefficients associated to the frequencies $f_i \leq f_{\max}$,

$$X = x + p_0 + \sqrt{2} \sum_{k=1}^{N/2} p_k \cos \left[\frac{\pi}{N} k \left(x + \frac{1}{2} \right) \right], \quad (4)$$

$$Y = y + q_0 + \sqrt{2} \sum_{k=1}^{N/2} q_k \cos \left[\frac{\pi}{N} k \left(y + \frac{1}{2} \right) \right]. \quad (5)$$

The use of the DCT implies that the N control points are equally spaced along the x -axis in the current frame.

4 Transformation Function Estimation

The parameters of the transformation functions in Eqs. (4) and (5) are estimated following a similar approach to that followed by Stevenson and Roorda [14]. A frame with minimal distortion is manually selected as the reference, while the other frame (current) is divided into strips, each of them with only a few lines along the direction of the fast scanner (y -axis). The actual number of lines on each strip must be manually selected, based on the SNR and the structure present on each data set. Then, the NCC between each strip and the reference frame is calculated, and its maximum within a certain region of interest (ROI) located. The dimensions and location of the ROI is determined by the maximum eye motion considered acceptable. The position of the maximum within the NCC matrix corresponds to the x - and y -shifts of the strip with respect to the reference frame. The definition of the NCC between the reference frame (R) and the current strip (S) used in this work is,

$$C_{R,S}(m,n) = \frac{\sum_{i,j} R(i,j) S(m+i,n+j)}{\sqrt{\sum_{i,j} R(i,j)^2 \sum_{p,q} S(p,q)^2}}, \quad (6)$$

where the sums are performed over pixels in the overlap area between the reference frame and the strip. The actual calculations are performed by taking advantage of the correlation theorem in the Fourier domain, a fast implementation of the discrete Fourier transform (DFT), and with adequate zero-padding to avoid periodization artifacts. The NCC definition in terms of the DFT can be written as

$$C_{R,S} = \frac{\text{IDFT} [\text{DFT} [R_p]^* \text{DFT} [S_p]]}{\sqrt{\text{IDFT} [\text{DFT} [P_R]^* \text{DFT} [S_p^2]] \text{IDFT} [\text{DFT} [R_p^2]^* \text{DFT} [P_S]]}}, \quad (7)$$

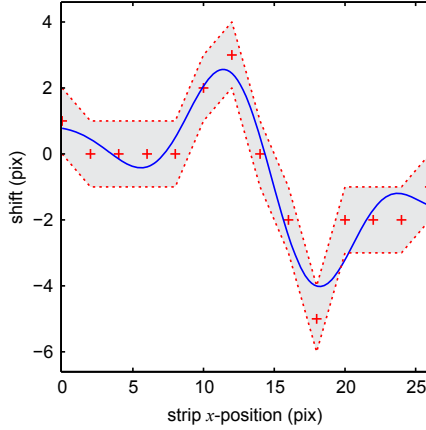


Fig. 1. Fitted ϵ_x model (blue line) of a sample set of current frame strip shifts (red crosses) using only the first $N/2$ coefficients returned by the DCT. The shaded area indicates the acceptable region, based on a shift uncertainty of 1 pixel. Note that although the fit might appear poor with respect to the data points, it is consistent with the uncertainty limits. In this example, each image strip consisted of two lines.

where IDFT denotes inverse DFT, $*$ indicates complex conjugate and R_p and S_p are the zero-padded reference frame and strip, respectively. P_R and P_S are zero-padded templates of the reference frame and strip, with unit value over the corresponding pixels. It is worth noting that the use of the NCC does not require any prior knowledge of the structures being imaged, and it uses all information in the image, as opposed to other algorithms with lower computational complexity [22]. In addition, the normalization used here makes the shift estimation robust to overall intensity fluctuations, which are very common in ophthalmic instruments, due to subject misalignment, tear film break up, etc. The NCC was implemented using the correlation theorem (Eq. 7) and the graphic processing units of CUDA-enabled graphics cards (Nvidia Corporation, Santa Clara, California, USA). The DFT was calculated using the CuFFT function. Figure 2 shows a plot of the performance of the NCC implemented using the CuFFT using single- and double-precision, against a double-precision Matlab implementation of Eq. 7 using the FFTW package (<http://www.fftw.org/>). Given current frame rates in ophthalmic instruments, the achieved performance allows real-time image registration, that produces motion signals that can be fed back to the scanning mirrors, to stabilize the imaging raster in the selected retinal location. There are two sources of error inherent to the shift estimation method. First, the use of the NCC on images with a finite number of pixels, which leads to a quantization of the shift values. This error can be reduced to an arbitrary level of accuracy by increasing the pixel density through interpolation, provided that computing power and memory requirements are not a limitation. The second source of uncertainty in the shift estimation comes from distortion within each strip. This error can be reduced by decreasing the number of lines in each strip, provided the information left

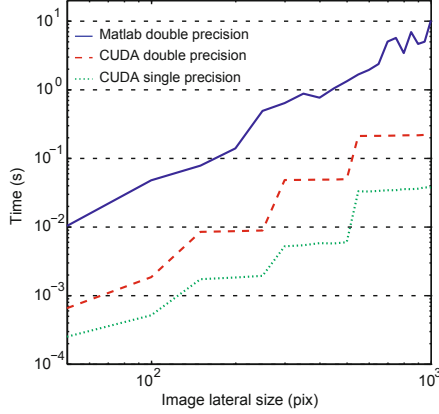


Fig. 2. Evaluation of the NCC implementation using CuFFT (CUDA) running on a GeForce GTX 285 graphics card from Nvidia for different (square) frame sizes. For reference, the performance of the FFTW in Matlab running on an Intel Xeon CPU E5430 using Microsoft Windows XP is provided.

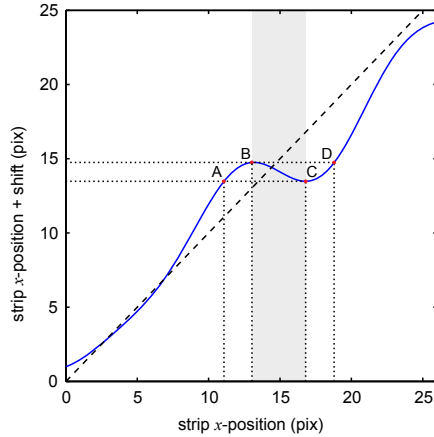


Fig. 3. Transformation model (blue line) corresponding to the ϵ_x from figure (II). The black dashed line shows the linear term of Eq. (4) that would correspond to the model in the absence of eye movement. The portion of the current frame corresponding to the decreasing part of the curve (shaded area) is not considered for the image registration.

within the strip is sufficient to correctly estimate the shifts. Once the x - and y -shifts of the (N) strips are estimated, they are used to fit the models in Eqs. (4) and (5). Note that the curve calculated using only $N/2$ cosine terms will in most cases not pass through all the control points (see figure I). If the sample density of the strip and reference frame are not increased when calculating the

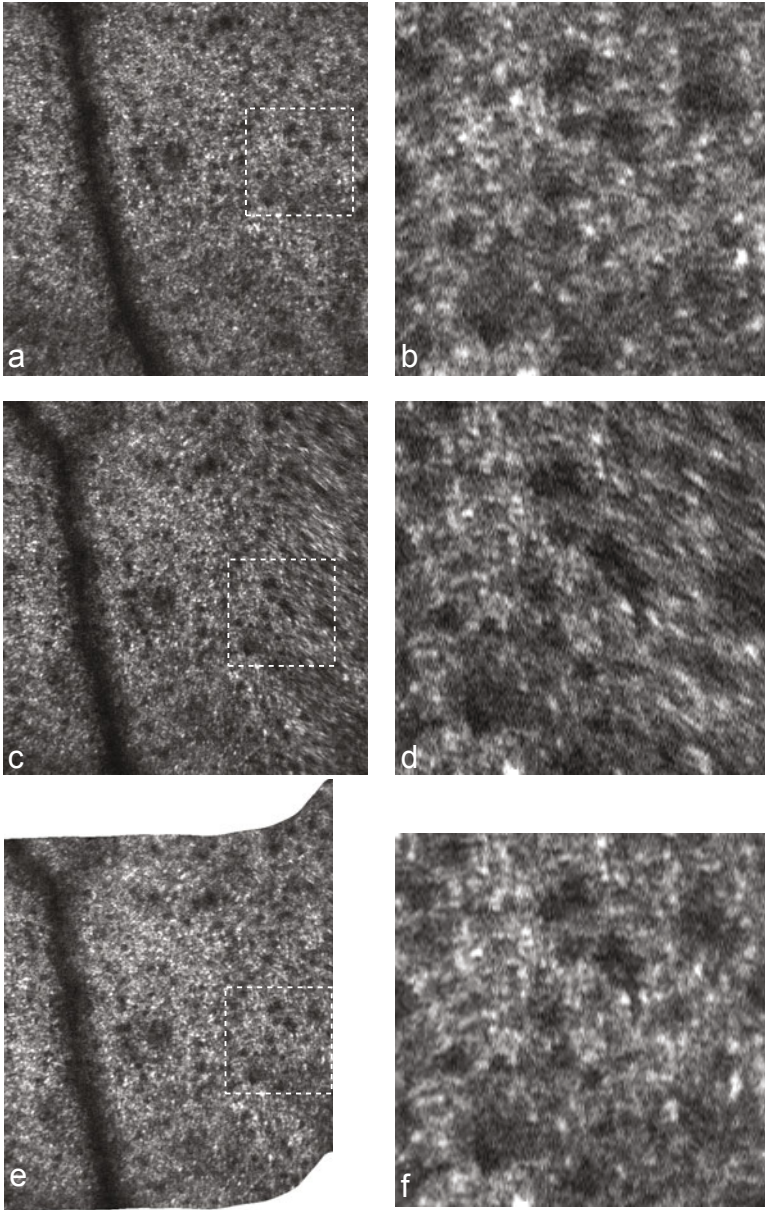


Fig. 4. Photoreceptor layer of a subject with poor fixation: (a) reference frame with minimal distortion, (c) a frame affected by eye motion and (e) the same frame after registration with the strip-based method. The images in panels (b), (d) and (f) show the regions indicated with the dashed squares in (a), (c), and (e) respectively.

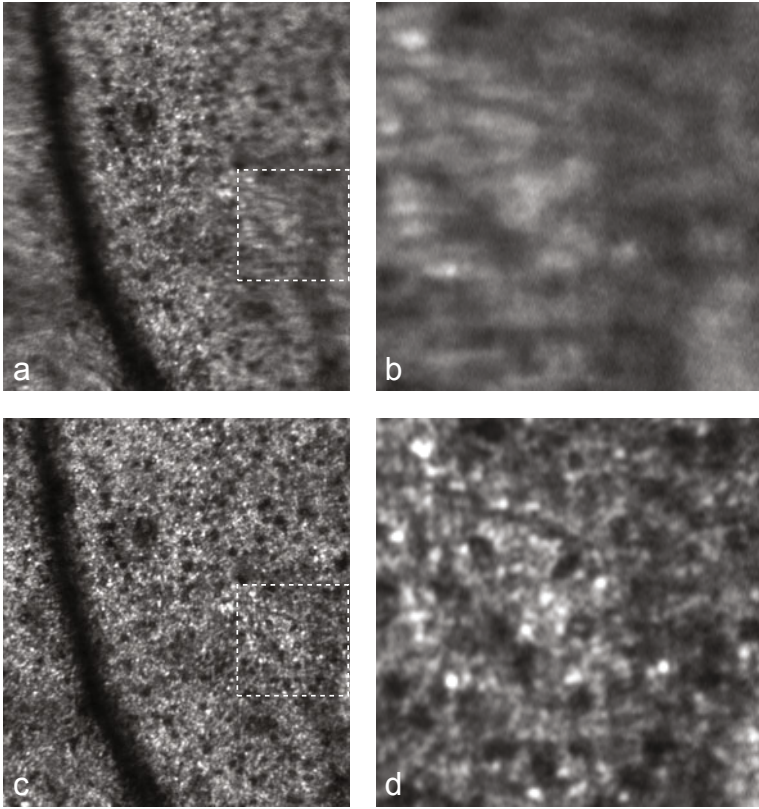


Fig. 5. Images that result from registering the sequence of 50 images that include those of Fig. 4 using: (a) the rigid registration and (b) the method proposed in this work. The latter method used 16 pixel-wide strips in a sequence of images of 708×688 pixels. The panels (b) and (d) show the regions indicated with the dashed squares in (a) and (c), respectively. Note the increased contrast and resolution achieved with the strip-based registration method, and the noise reduction with respect to figure 4.

NCC, then the minimum uncertainty is due to shift quantization into integer values (± 0.5 pixels). As mentioned before, another source of error is the image distortion within the strip. The effect of this source of error is difficult to quantify, as it varies with the structure being imaged, the amplitude, and speed of eye movement. A metric of distortion within a strip is the NCC value associated with the estimated shifts. These NCC values can be compared against a threshold in order to accept or discard strips.

Once the transformation functions are estimated from the strip shifts, the resulting curve from the x -shift has to be split into strictly monotonic intervals. When the curve is not strictly increasing, the same retinal patch is imaged multiple times in a single current frame. For example, according to the curve in Fig. (3), the section of the current frame corresponding to the intervals AB, BC

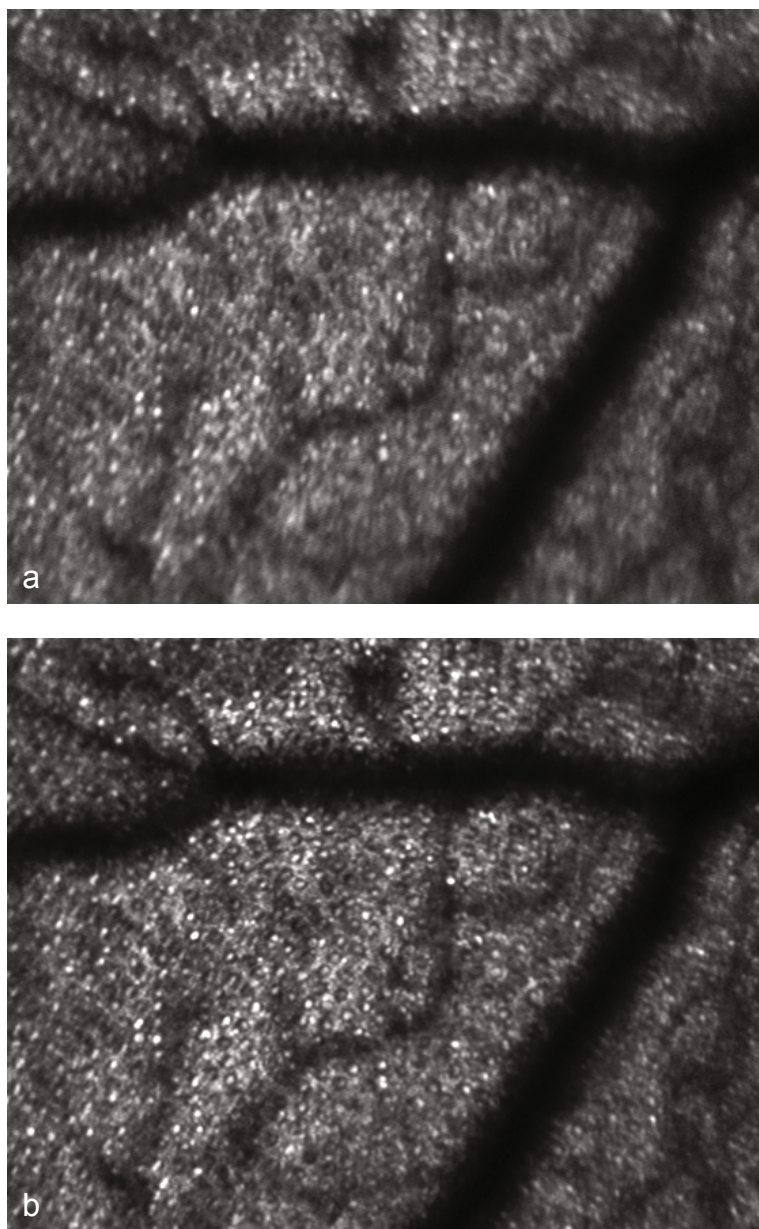


Fig. 6. Output image showing the photoreceptor mosaic of a healthy human subject with stable fixation, after registering a sequence of 500 images acquired using a reflectance AOSLO using: (a) rigid registration and (b) the method proposed in this work. Note the contrast and resolution increase achieved with the strip-based method.

and CD would all show the same retinal patch. The BC segment indicates retinal movement in the direction of the slow scanner. In practice, the corresponding portion of the image would be inverted and too distorted to be useful. There is no need to verify that the y -transformation function is monotonic because of the high speed of the fast scanner with respect to eye motion.

Finally, each monotonic interval in the x -transformation function and the corresponding interval in the y -transformation function are used to calculate the values of the registered image over the grid of pixels of the reference frame, using a bilinear interpolation. In this way, the reference and the registered images can be summed to produce images with higher SNR.

Figures 4, 5 and 6 illustrate how the proposed strip-based method works and how it performs with respect to a simple rigid translation. Figures 4, and 5 show the retina of a human subject suffering from blue cone monochromacy, a condition that results in very poor visual acuity and fixation. The panels in figure 4 are single frames taken from a sequence of 50 frames, while those in figure 5 correspond to the averaged registered sequence. For comparison, registered images from a subject with no eye conditions and good fixation are provided in Fig. 6.

5 Summary and Discussion

We have shown that when the scanning in ophthalmic instruments is sufficiently fast with respect to the speed of involuntary eye movement, the mapping functions become the addition of a linear term and a single variable function. Then, based on experimental data on eye movement amplitude and speed in the fixating eye, minimum sampling frequencies for these instruments were discussed. Assuming finite bandwidth of the involuntary eye motion, a method for estimating the transformation functions was presented. The proposed registration method can be used to improve the signal-to-noise ratio (SNR) of high-resolution reflectance, single-photon fluorescence and phase images from the live retina. Currently, the only part of the registration process that requires significant user input is the selection of the reference frame, which is also necessary for rigid registration or any other registration method. Finally, two examples illustrating the dramatic image quality improvement provided by the proposed method with respect to rigid translation were presented.

Acknowledgements

Alfredo Dubra-Suarez, Ph.D., holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. This research was partially supported by the National Institute for Health, Bethesda, Maryland through the grants BRP-EY014375 and 5 K23 EY016700.

References

1. Liang, J., Williams, D.R., Miller, D.T.: Supernormal vision and high-resolution retinal imaging through adaptive optics. *J. Opt. Soc. Am. A* 14(11), 2884–2892 (1997)

2. Rha, J., Jonnal, R.S., Thorn, K.E., Qu, J., Zhang, Y., Miller, D.T.: Adaptive optics flood-illumination camera for high speed retinal imaging. *Opt. Exp.* 14(10), 4552–4569 (2006)
3. Hermann, B., Fernandez, E.J., Unterhuber, A., Sattmann, H., Fercher, A.F., Drexler, W., Prieto, P.M., Artal, P.: Adaptive-optics ultrahigh-resolution optical coherence tomography. *Opt. Lett.* 29(18), 2142–2144 (2004)
4. Zawadzki, R., Jones, S., Olivier, S., Zhao, M., Bower, B., Izatt, J., Choi, S., Laut, S., Werner, J.: Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal in vivo imaging. *Opt. Exp.* 13(21), 8532–8546 (2005)
5. Fernandez, E.J., Povazay, B., Hermann, B., Unterhuber, A., Sattmann, H., Prieto, P.M., Leitgeb, R., Ahnelt, P., Artal, P., Drexler, W.: Three-dimensional adaptive optics ultrahigh-resolution optical coherence tomography using a liquid crystal spatial light modulator. *Vision Res.* 14(20), 8900–8917 (2006)
6. Zhang, Y., Rha, J., Jonnal, R., Miller, D.: Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina. *Opt. Exp.* 13(12), 4792–4811 (2005)
7. Merino, D., Dainty, C., Bradu, A., Podoleanu, A.G.: Adaptive optics enhanced simultaneous en-face optical coherence tomography and scanning laser ophthalmoscopy. *Opt. Exp.* 14(8), 3345–3353 (2006)
8. Bigelow, C.E., Iftimia, N.V., Ferguson, R.D., Ustun, T.E., Bloom, B., Hammer, D.X.: Compact multimodal adaptive-optics spectral-domain optical coherence tomography instrument for retinal imaging. *J. Opt. Soc. Am. A* 24(5), 1327–1336 (2007)
9. Roorda, A., Romero-Borja, F., Donnelly III, W.J., Queener, H., Hebert, T.J., Campbell, M.C.W.: Adaptive optics scanning laser ophthalmoscopy. *Opt. Exp.* 10(9), 405–412 (2002)
10. Gray, D.C., Merigan, W., Wolfing, J.I., Gee, B.P., Porter, J., Dubra, A., Twietmeyer, T.H., Ahamd, K., Tumber, R., Reinholz, F., Williams, D.R.: In vivo fluorescence imaging of primate retinal ganglion cells and retinal pigment epithelial cells. *Opt. Exp.* 14(16), 7144–7158 (2006)
11. Burns, S.A., Tumber, R., Elsner, A.E., Ferguson, D., Hammer, D.X.: Large-field-of-view, modular, stabilized, adaptive-optics-based scanning laser ophthalmoscope. *J. Opt. Soc. Am. A* 24(5), 1313–1326 (2007)
12. Morgan, J.I.W., Dubra, A., Wolfe, R., Merigan, W.H., Williams, D.R.: In vivo autofluorescence imaging of the human and macaque retinal pigment epithelial cell mosaic. *Invest. Ophthalm. Vis. Sci.* 50(3), 1350–1359 (2009)
13. American National Standard for safe use of lasers (ANSI Z136.1), Laser Institute of America, Orlando, Florida, USA (2007)
14. Stevenson, S.B., Roorda, A.: Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy. In: *Proc. SPIE*, vol. 5688A, pp. 145–151 (2005)
15. Hart Jr., W.H.: *Adler's physiology of the eye: clinical application*, 9th edn., Mosby-Year Book Inc., 11830 Westline Industrial Drive, St. Louis, Missouri 63146 (1992)
16. Riggs, L.A., Armington, J.C., Ratliff, F.: Motions of the retinal image during fixation. *J. Opt. Soc. Am.* 44(4), 315–321 (1954)
17. Eizenman, M., Hallett, P.E., Frecker, R.C.: Power spectra for ocular drift and tremor. *Vision Res.* 25(11), 1635–1640 (1985)

18. Ditchburn, R.W., Ginsborg, B.L.: Involuntary eye movements during fixation. *J. Physiol.* 119, 1–17 (1953)
19. Whittaker, S.G., Budd, J., Cummings, R.W.: Eccentric fixation with macular scotoma. *Invest. Ophth. Vis. Sci.* 29(2), 268–278 (1988)
20. Smith, G., Atchison, D.A.: *The eye and visual optical instrumentation*, 1st edn. Cambridge University Press, Cambridge (1997)
21. Charman, N.: *Handbook of optics: Vision and vision optics*. In: Bass, M. (ed.) *Optics of the Eye*, ch. 1, 3rd edn., vol. III. Mc Graw Hill, New York (2009)
22. Arathorn, D.W., Yang, Q., Vogel, C.R., Zhang, Y., Tiruveedhula, P., Roorda, A.: Retinally stabilized cone-targeted stimulus delivery. *Opt. Exp.* 15(21), 13731–13744 (2007)