

PAPER • OPEN ACCESS

Precision targeting for retinal motion extraction using cross-correlation with a high speed line scanning ophthalmoscope

To cite this article: Yi He *et al* 2015 *J. Opt.* **17** 125303

View the [article online](#) for updates and enhancements.

You may also like

- [Performance measurements of an infrared digital scanning laser ophthalmoscope](#)
A Manivannan, P F Sharp and J V Forrester
- [Extraction of ultra-high frequency retinal motions with a line scanning quasi-confocal ophthalmoscope](#)
Yi He, Ling Wei, Zhibin Wang et al.
- [The scanning laser ophthalmoscope](#)
P F Sharp and A Manivannan

Precision targeting for retinal motion extraction using cross-correlation with a high speed line scanning ophthalmoscope

Yi He^{1,2}, Ling Wei^{1,2}, Zhibin Wang^{1,2,3}, Jinsheng Yang^{1,2}, Li Xiqi^{1,2}, Guohua Shi^{1,2} and Yudong Zhang²

¹The Key Laboratory on Adaptive Optics, Chinese Academy of Sciences, Chengdu 610209, People's Republic of China

²Institute of Optics and Electronic, Chinese Academy of Sciences, Chengdu 610209, People's Republic of China

³Graduate School of Chinese Academy of Sciences, Beijing 100039, People's Republic of China

E-mail: heyi_job@126.com and ioe_eye@126.com

Received 20 February 2015, revised 9 May 2015

Accepted for publication 1 June 2015

Published 5 November 2015



Abstract

Evaluations to quantify the precise targeting of template features and to select template sizes for retinal motion extraction were carried out using cross-correlation with a high speed line scanning ophthalmoscope (LSO) capable of 160 frames per second. The optimal template targeting was located on a retinal vessel pattern with vessel bifurcation or vessel features occupying approximately eighty percent of the template area preferred. The optimal template size for this LSO system was 80×80 pixels and it was able to extract retinal motion up to 300 deg s^{-1} at a speed of 30 Hz. Although the optimized template size was a compromise between having enough image data on the retinal features to make matches reliably and have good temporal resolution, the optimal targeting of the template location and size described here was appropriate and effective in extracting retinal motion. In addition, the determination of cross-correlation templates could be applied to other images having similar properties; i.e., relatively small features of distinct gray levels on an otherwise fairly uniform background.

Online supplementary data available from stacks.iop.org/JOPT/17/125303/mmedia

Keywords: vision—eye movements, confocal microscopy, retina scanning, ophthalmic optics devices

1. Introduction

The human eye continually moves even when we steadily fix our gaze on an object [1]. These motions include high frequency tremors, rapid microsaccades, and slower drifts with an amplitude of several arc seconds to several arc minutes and a frequency of 10–100 Hz [2–4]. These movements are too small to be seen with the naked eye or mid-level eye movement monitors, but they introduce significant inter-frame

distortions in retina images captured from fundus camera [5–7]. What is worse, they produce intra-frame and inter-frame distortions, and can warp images during eye fundus imaging examinations with high resolution systems such as optical coherence tomography [8–10], and confocal scanning laser ophthalmoscopy (SLO [11–13]). Therefore, image artifacts introduced by these movements severely impact accurate visualization of the retina, and hinder mathematical measurements of retinal microstructures [14]. For example, retinal motion lowers the quality of images, making it difficult to quantify measurements obtained by high resolution retinal imaging systems such as retinal thickness mapping (retinal topography) [15], retinal cell density mapping and so on [16].



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

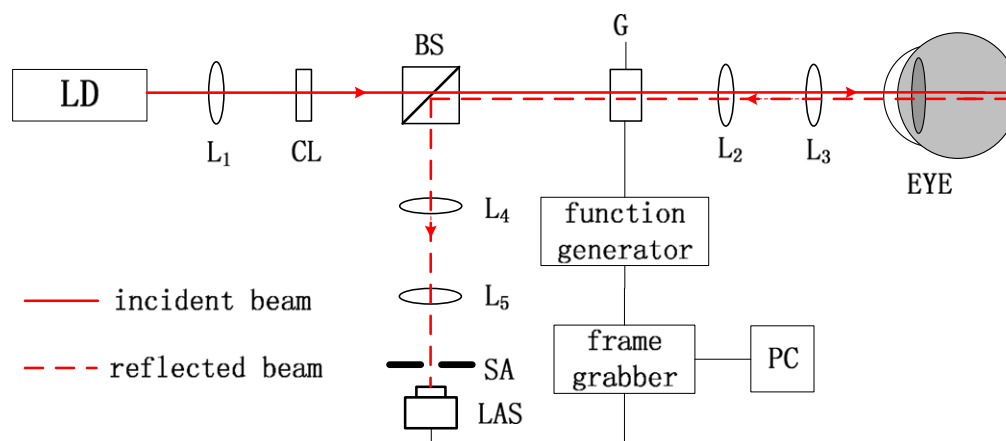


Figure 1. Schematic of the LSO system. LD, laser diode; BS, beam splitter; CL, cylindrical lens; L1–L5, lenses; SA, slit aperture; G, galvanometer-driven scanner; LAS, linear array sensor.

For many decades, cross-correlation has been the most widely used method for retinal motion extraction. The earliest ophthalmoscopic system for retinal motion measurement was implemented by Cornsweet, who reported a resolution of ten arc seconds with a system that tracked a blood vessel using an oscilloscope and photomultiplier tube [17]. O'Connor applied a frame-to-frame cross-correlation to determine inter-frame retinal movements in SLO images [18], and then patch-based cross-correlation methods were proposed to estimate retinal movements in the SLO data, which required only a sequence of frames of scan data and estimated intra-frame motions [19]. Vogel *et al* applied a map-seeking circuit algorithm to compute correlations to estimate retinal motion vectors and account for more general motions [20]. These approaches have been shown to accurately estimate intra-frame retinal motions that are close to 1 KHz in offline analysis; however, only drift motions at slow speeds could be estimated while ultra-high frequency inter-frame distortions such as tremors or rapid microsaccades were unable to be estimated clearly [21]. Since the frame rate cannot be accelerated to overcome intra-frame retinal motions, these approaches led to failure in estimations due to displacements in correlation templates in sequential frames [22].

Therefore, the most efficient approach to estimate ultra-high frequency retinal movements is to accelerate the imaging speed. Increasing image acquisition speed using a high speed line scanning ophthalmoscope (LSO) with a frame rate of capable of 160 frames per second (fps) [23, 24] eliminates motion artifacts within frames and reduces feature displacement between frames, and the absence of intra-frame motion gratifies extraction of retinal motion through cross-correlations from inter-frame distortions [25].

However, cross-correlation involves computing image intensities to maximize similarities between a pair of discrete rectangular images so that templates of the cross-correlation are related to detection accuracy. In brief, the template location and size related to different retinal features influences the precision of cross-correlation calculations. As the template locates different features of the retina, the contrast and gray level difference between the template features and the retinal

background may be distinct or uniform, which may be a reliable or erroneous cross-correlation [26]. When the template contains optimum features of the retina, the template size is associated with efficiency of the cross-correlation. As the template size increases, the precision of cross-correlation computations increase, but only to a relatively stable point. Beyond this point, the cross-correlation will increase slowly and infrequently. Moreover, the anticipated high velocity of retinal motion is proportional to the template size [27]. However, given that the precision of cross-correlations is directly proportional to the increase of the template size to a certain extent, one cannot neglect the computational complexity with a large template size. The cross-correlation technique is computationally intensive and requires a lengthy calculation period despite the use of fast mainframe computers [28], and a smaller template size makes the cross-correlation process even faster. Consequently, the optimal template size will be a compromise between having enough retinal feature image data to match reliably and having good temporal resolution [29].

Design of a precise template targeting methodology, including template location and size, is critical to carry out cross-correlations to extract retinal motion. In this study, we extended previous studies by rigorously evaluating the performance of cross-correlations to extract retinal motion. Correlation templates of different retinal features were used to determine the optimal template for retinal structures, and several criteria were compared to determine the optimum template size of the cross-correlation as it was applied to extract retinal motion. Evaluations quantified the precise targeting of template features and template sizes for retinal motion extraction using cross-correlations.

2. Materials and method

2.1. Experimented configuration

A detailed description of the LSO system can be found in Yi He *et al* [30] and the optical layout can be seen in figure 1. A

laser diode (LD) with a central wavelength of 780 nm was collimated through a convex lens (L1) and spread in one dimension with a cylindrical lens (CL). The line beam was focused onto the retina with a scanning lens (L2), and the retina was scanned with a galvanometer-driven mirror (G, 6230H, Cambridge Corporation). The CL and the galvanometer, as shown in figure 1, were used as a line beam scanner, which ultimately converted the point light source onto a focused line beam. The back-scattered light from the retina passed through the same lens and was descanned by the galvanometer through the beam splitter (BS) and focused onto the linear array sensor (LAS, AVIVA M2 CL 1014, ATMEL) with a detector objective lens (L3). Meanwhile, a slit aperture (SA) conjugated to the retinal plane was placed close to the LAS to reject the majority of the back-scattered light from the adjacent voxels along the scanned line.

As the line acquisition rate of the line-CCD was 53 kHz, the image acquisition speed could reach up to 160 fps for 800×400 pixels, or $2400 \times 1200 \mu\text{m}$ on a retinal scale. The whole length of the linear array sensor is 1024 elements, but the curve shape of the retina produces some dark fields on the margin, and the middle pixels of 800 elements are useful to collect light. According to theoretical optical performance characterized by ZEMAX (ZEMAX Development Corporation, USA), the LSO system achieves diffraction-limited performance through the entire scanning angle (3 mm non-mydratic pupil, 8° on retina plane), and the transverse Airy disk diameter referenced to the retina is $10 \mu\text{m}$. The retinal conjugate was magnified by 4.8 on the linear array (also the confocal SA), and a $50 \mu\text{m}$ SA ($\sim 10.05 \mu\text{m}$ on the retina) was chosen, which means it was tightly confocal (~ 1.05 times the Airy disc). Accordingly, an estimated optical resolution was approximately $10 \mu\text{m}$, which was limited by the aberrations of the human eye.

Prior to examination, informed consent was obtained from all of the subjects. An important informed consent was also obtained for the power of the light beam. The incident power at the cornea for the extended imaging beam was $500 \mu\text{W}$. This exposure level is considered safe according to ANSI standards [ANSI Z136.1-2000] for several hours of continuous intra-beam viewing. For our experiments, the line illumination would allow for even higher safe exposure levels, since the power was roughly 50 times below the ANSI maximum permissible exposure levels for the human eye.

The LSO system was tested on human volunteers aged 20 to 35 years with healthy eyes. In order to minimize the characteristics of retinal movements, the chosen subjects were typically highly practiced observers with extensive knowledge on controlling eye movement behavior. Therefore, it can be assumed that they exercised great efforts to maintain a steady gaze. In addition, a chin rest and forehead rest were used to stabilize head movements, and the subjects were asked to fixate on a bright green target with the fellow eye.

2.2. Preliminary human subject test plan

For the high speed imaging sequences of our LSO system, as discussed above, there were almost no remnant intra-frame

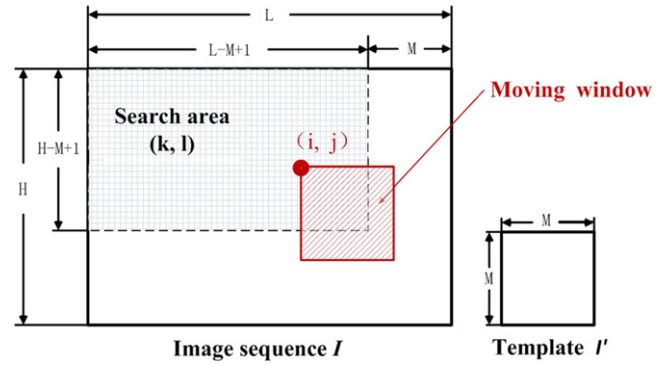


Figure 2. The cross-correlation algorithm to extract retinal motions from LSO frames.

distortions in the single frame and the image sequences barely differed in magnification and rotation. In addition, fixation made the rotation of retinal motion negligible [31]. Consequently, the determination of similarity between sequential images yielded the best translational fit. The most widely used method for similarity detection is cross-correlation.

The details of retinal motion recovery with this system have been previously reported [32]. Briefly, an image is used as a fixed reference frame (usually taken to be the first frame), and a template of image features is selected on the reference image. As illustrated in figure 2, the whole window of sequences is the search area. The template window is then shifted across the search area, and a similarity measure is calculated. The point of maximal similarity is designed as the match position, where the features in the template are aligned with the corresponding features in the search area of image sequences to determine the translational motions.

An exhaustive search in a cross-correlation technique is defined as a pixel-by-pixel scan of sequential images to find the match coordinates, which is computationally intensive and may require a lengthy calculation period. For this reason, instead of a standard cross-correlation to calculate similarity detection, some novel techniques have employed an enhancing cross-correlation measure, and a normalized cross-correlation was subsequently used to correct for differences in illumination, which was defined as:

$$\begin{aligned} \text{corr}(I, I')_{k,l} &= \frac{\sum_i \sum_j (I(i+k, j+l) - E(I)) \times (I'(i, j) - E(I'))}{\sqrt{\sum_i \sum_j (I(i+k, j+l) - E(I))^2 \times \sum_i \sum_j (I'(i, j) - E(I'))^2}}, \end{aligned} \quad (1)$$

where I and I' are the reference image and sequential image, respectively, and $E(\cdot)$ is the expectation of images, k and l denotes the position of the moving windows. By finding the indices k and l , which maximize the cross-correlation $\text{corr}(I, I')_{k,l}$, the (k, l) displacements of the new frame with respect to the reference frame were determined to be a measure of the relative motion of the retina within the specific area.

In order to improve the reliability and accuracy of the cross-correlation approach, we added a time-saving procedure. A modified cross-correlation was developed and programmed. One advanced modification was that a 'limited exhaustive search' routine was used in the cross correlation method in which the search area was restricted to pixels within the radius of the anticipated maximum retinal movement between frames by a 'look-ahead' strategy based on motion predicted from the previous frame.

The cross-correlation approach is a computationally efficient alternative to other motion estimation methods, allowing estimates of retinal motion that allow comparisons of imaging frame rates. In addition, it is not sensitive to image details, so image features of retinal structures including photoreceptors, feeder vessels, nerve fibers, and other cells can be extracted to measure retinal motion.

3. Results and discussion

Retinal motion was extracted from the raw LSO videos using a cross-correlation procedure. All frames in the video were compared to a single reference frame (taken as the first frame) in order to determine how the retina had moved during the course of imaging. Note that there are many features of the retina to be selected as a cross-correlation template to extract motions, and marking of different features, such as blood vessels, photoreceptors, and other cells may affect the precision of motion retrieval.

In order to determine precision targeting for retinal motion extraction, motion estimations based on manual marking of different retinal features was performed, and we present. Once precision targeting of retinal features was executed, the other important consideration was the size of the template for motion extraction in the cross-correlation approach. When the template size was too small, the accuracy of motion estimation decreased severely. A much larger cross-correlation template impeded the speed of the motion extraction procedure. In section 3.2, we present the experimental results for different template sizes, which determined the effectiveness and high-speed characteristics of the motion extraction scheme.

3.1. Tracking templates with different features

The video clip in figure 3 shows an 877 frame sequence (5.48 s) of LSO data. Feature displacements occurred at all times due to inter-frame motions. One single frame is shown in figure 3 to demonstrate retinal features that are potential template targets for motion extraction using the cross-correlation method.

The particular features in the image included blood vessels of different scales, parallel vessels and bifurcations, avascular areas, and the optic nerve head. An arbitrary retinal feature can be selected as a template in the cross-correlation tracking method, and templates of different retinal features do not affect the computation speed. Recall that the cross-correlation approach employs computing image intensities to

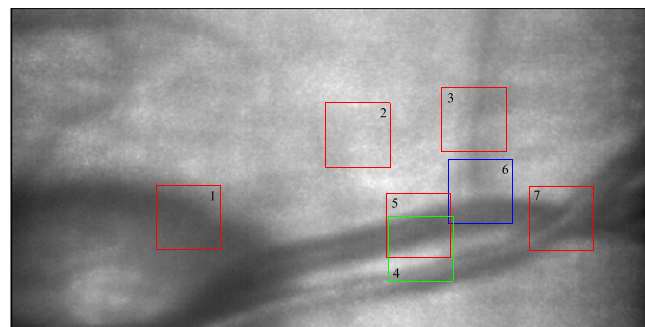


Figure 3. Different features as potential template targets are exhibited in a single frame. The color frames of No. 1–7 mark templates of different retinal features: optic nerve head, avascular area, small vessel, parallel vessels, large vessel, vessel bifurcation in a small scale, and vessel bifurcation in a large scale in sequence. The video consisted of 877 image frames for a duration of 5.48 s, and the image size was 800×400 pixels, or 8×4 degrees, or $2400 \times 1200 \mu\text{m}$ on a retinal scale. (See supplementary data, available at stacks.iop.org/JOPT/17/125303/mmedia.)

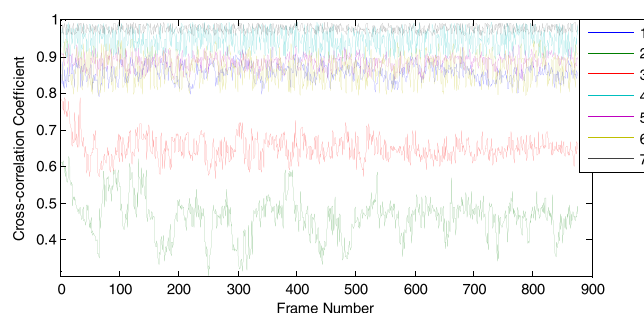


Figure 4. The cross-correlation coefficients obtained from different templates in cross-correlations.

maximize similarities between a pair of discrete rectangular images, so that templates of different features are related to cross-correlation detection accuracy.

We chose seven distinct locations in the retina image as templates (labeled in figure 3), and retinal motion extraction experiments were executed using the cross-correlation scheme. Note that retinal motion with large velocities diminished the visibility of retinal features in the search templates. Accordingly, the template had variable size flexibility if the predicted template location was missing in the field of view due to large motions. In this case, the location and size of the template was preserved, and the missing portion of the template was cut off; only the remnant template was used to cross-correlate for extracting retinal motion. This variety of templates assured the cross-correlation procedure was successful and reduced false extractions.

Note that the cross-correlation coefficient $\text{corr}(\mathbf{I}, \mathbf{I}')_{k,l}$ presented in equation (1) defined the precision of retinal motion extraction. The closer $\text{corr}(\mathbf{I}, \mathbf{I}')_{k,l}$ was to 1, the more similar the sequence images became. Figure 4 shows the cross-correlation coefficient $\text{corr}(\mathbf{I}, \mathbf{I}')_{k,l}$ obtained from different templates. The positive direction in figure 4 corresponds to a higher precision of retinal motion estimates. The color labels in figure 4 correspond to template locations of distinct features in figure 3. Note that we have used simple

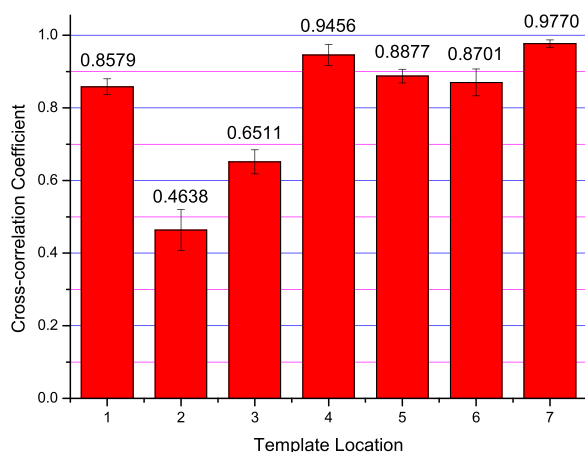


Figure 5. Effectiveness of cross-correlations for templates employing different retinal features. The cross-correlation coefficient was determined for all seven template locations in the same imaging session and were averaged across all frames in the video. Error bars denote $\pm$ one standard deviation of coefficient across all frames in the video.

linear interpolation to fill in the missing coefficient data due to the discrete scan mode and due to a lack of recorded data during large motions in which the template location left the field of view.

In figure 4, one can clearly see some obvious distributions for cross-correlation coefficients. Cross-correlation coefficients in templates 2 and 3 were smaller than 0.8 with a relatively flat peak of cross-correlation values. Middle-level accuracies in the cross-correlation located in templates 1, 5, and 6 had a coefficient between 0.8 and 0.9. Due to the comparatively distinct features of vessels, a much higher coefficient for the cross-correlation was observed in templates 4 and 7. Occasionally, the cross-correlation of large parallel vessels in template 4 performed even more exactly than the vessel bifurcation in template 7. Overall, vessel bifurcation performed significantly better than parallel vessels. The difference was in the reliability of the cross-correlation at templates using different retinal features.

In order to manifest cross-correlation extractions clearly, the mean values of cross-correlation coefficients of different templates are shown in figure 5, and the standard error of the mean (SEM) is also plotted in this figure. In the plot, each bar has two numbers; i.e., an average (plotted histogram) and a SEM (error bar). The cross-correlation locations for template 2 with the lowest accuracy had a coefficient of 0.4638. This was due to template 2 having no particular features and a fairly uniform background, and the cross-correlation process performed a relative fluctuation of the largest SEM. Template 3 had a small vessel with an indistinct contour, so the coefficient was slightly higher than template 2, but was still below 0.7. Templates 1, 5, and 6 occupied a large vascular pattern of the retina, and cross-correlation coefficients were as high as 0.8. A higher precision of cross-correlation occurred in templates 4 and 7, with a coefficient greater than 0.9 and a much smaller SEM value. Allowing for stability in the cross-correlation procedure with a small SEM value, features in

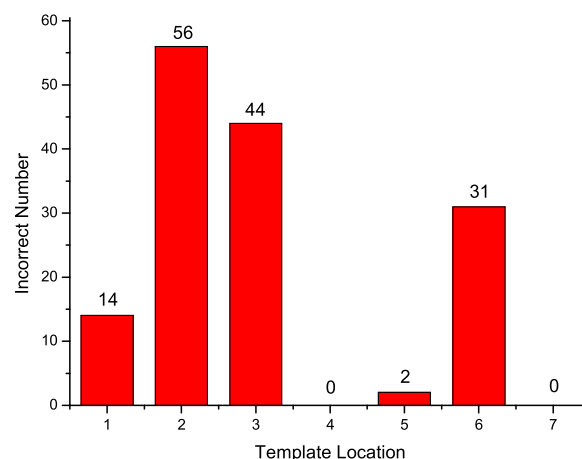


Figure 6. Numbers of erroneous cross-correlations with templates based on different retinal features.

template 7 were more efficient cues for extracting retinal motion with the cross correlation method.

Normally, one would expect better detection with higher cross-correlation coefficients, because this showed the testing was more accurate. However, an unacceptable, erroneous correlation occurred many times when higher values of cross-correlation coefficients were calculated. Since the cross-correlation computed the coefficient in the searched area from sequential images and produced a maximal similarity value, higher values of cross-correlation coefficients did not signify true correlation detections, especially when there was a large SEM value. Therefore, it was necessary to pick out incorrect similarity operations to generate an overall inspection of the cross-correlation extraction.

All the template experiments were examined to sum the number of incorrect cross-correlation operations. As one general criterion for correlation, once the shift position of the template in this calculation exceeded one pixel compared to the previous measurement, the cross-correlation estimation was determined to be one incorrect cross-correlation operation. Figure 6 shows the number of incorrect cross-correlations for all template experiments. With the smallest coefficient and the largest SEM, the incorrect number for template 2 was the largest value; i.e., 56, implying the lowest precision correlation computation. The incorrect number for template 3; i.e., 44, was a little smaller than template 2, and this implied an unacceptable cross-correlation experiment. Although the mean values of coefficients for templates 1, 5 and 6 were nearly equivalent (shown in figure 6), the incorrect numbers were not the same (shown in figure 7). The large SEM value for template 6 implied an unstable cross-correlation, and produced an incorrect number of 31. The incorrect numbers for templates 1 and 5 were decreased greatly, which are in accordance with the SEM values shown in figures 4 and 5. There were no incorrect cross-correlation operations with templates 4 and 7, and these two templates of retinal features are potentially the best targets for retinal motion extraction with the cross-correlation procedure.

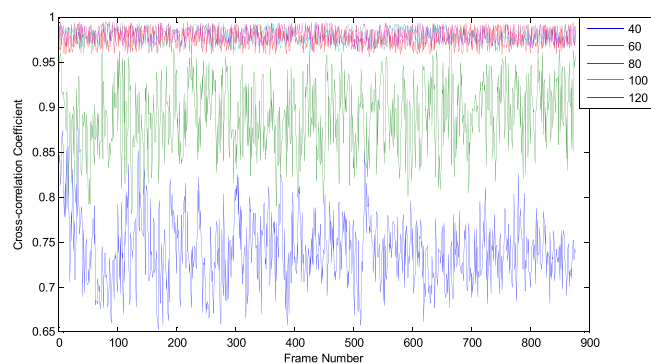


Figure 7. The cross-correlation coefficients obtained with different template sizes.

In figure 6, only templates 4 and 7 showed no incorrect extraction of retinal motion. In accordance with the large coefficient values and small variances, shown in figures 4 and 5. Therefore, we can say that the template 7 is the optimal targeting location for the cross-correlation method, accompanied by the largest cross-correlation coefficient and the smallest SEM value.

The extraction of retinal motion using template 4 had a relatively high coefficient value, but the large SEM value indicated that the cross-correlation procedure fluctuated, and the correlation matching calculation had weak stability. Once the image quality decreased, features of template 4 had reduced correlation accuracy for the estimation of retinal motion.

Consequently, we can draw a clear conclusion. The template employing distinct retinal vessels had the highest accuracy and stability, and characteristics of blood features are the best potential templates in the cross-correlation algorithm. The vascular pattern of the retina should be used as the main template cue for cross-correlation extraction of retinal motion. The essential reason is that the retinal background is fairly uniform, and vessels are shown very distinctly. Ordinarily, a variety of vessel bifurcation structures appear in retinal imaging sessions. As a result, vessel bifurcation or vessel features occupying approximately eighty percent of the template area, which simultaneously has a non-uniform intensity distribution, are the optimal targets in cross-correlation templates for retinal motion estimations.

3.2. Determination of the template size

Cross-correlation has been shown to be reliable for extracting retinal motion, but it is computationally intensive. Modifications were made to produce a faster algorithm, and we made some new attempts to speed the calculations, as shown in section 2.2.

Considering computer capabilities today, there was no need for a special design and allocation of computer memory, and cross-correlation matching was quickly realized on a personal computer. While the cross-correlation was used for extraction of retinal high-frequency motion, several conditions needed to be met: (1) an execution speed for correlation computations is not much lower than the imaging speed, at

least in real time scales; and (2) tracking of all retinal motion components, especially for motions at high speeds and large amplitudes that are hard to extract. All the requirements were related to the template size of the correlation similarity.

Normally, a smaller template size reduces the correlation computations, and is computationally more effective; on the other hand, a smaller template includes fewer features in the calculation of correlation similarity, leading to a decreased accuracy of detection and, what is worse, special retinal motion components cannot be extracted completely. In our study, we determined how the template size affects the extraction of retinal motion using a cross-correlation method.

Note that we chose vessel bifurcation as the template location, labeled as 7 in figure 3, and we continued to execute the cross-correlation for the same LSO data as in figure 3. Each test was carried out at a number of template size settings: 40×40 pixels, 60×60 pixels, 80×80 pixels, 100×100 pixels, and 120×120 pixels. For each experiment, we recorded the performance of cross-correlations and completion times.

In the same manner, we plotted the cross-correlation coefficients obtained using different template sizes in figure 7. One can clearly see obvious distributions for cross-correlation coefficients shown in figure 7. When the template size was 40×40 pixels, the cross-correlation coefficients were smaller than 0.8, and with a relatively fluctuating cross-correlation procedure. It was more accuracy in the correlation experiment with 60×60 pixels for the template size, and the coefficient reached 0.9. When the template size increased to 80×80 pixels, and even larger sizes, the coefficients were close to 1. It was difficult to distinguish the results from these three sets, shown in figure 7. Along with the increase in template size, up to 80×80 pixels, the correlation coefficients increased, mainly because the template contained enough features to produce high accuracy cross-correlation similarity management.

The template located on the vessel bifurcation was labeled as No.7 in figure 3.

A statistical analysis of the coefficients shown in figure 7 was performed and the mean values and SEM values are shown in figure 8. The smallest template size of the cross-correlation produced a coefficient of 0.74 with a SEM of 0.38. A larger template of 60×60 pixels increased the precision of the cross-correlation, with a coefficient of 0.89, and the SEM was 0.39, which means there was a lack of stability for this template size. When the template size increased to 80×80 pixels, the coefficient increased to 0.977 with a small SEM of 0.010, signifying an improvement in stability. When the template size was larger than 80×80 pixels, the coefficients remained at 0.98 and with a SEM slightly less than 0.01.

Using the identical criteria for numbers of incorrect cross-correlations discussed above, the incorrect correlation performances are summarized in figure 9. As shown in figure 9, the larger the template size was, the lower the number of incorrect cross-correlations would be. There were 28 times incorrect cross-correlation operations when the template size was 40×40 pixels, signifying an unreliable correlation experiment due to lack of matching data. When

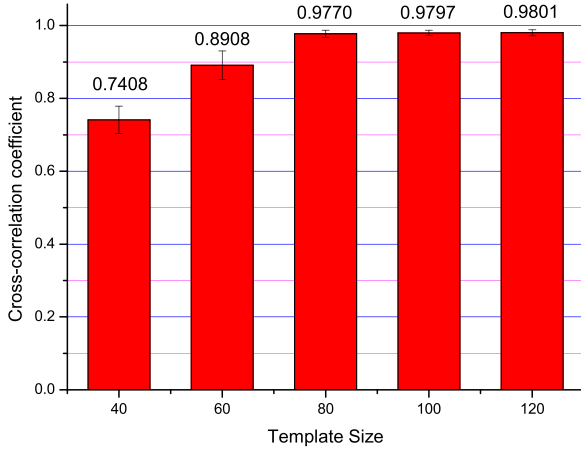


Figure 8. Plot of cross-correlation coefficient versus template size. Cross-correlation coefficients shown as mean values across all the frames in the video. Error bars denote $\pm$ one standard deviation of coefficient across all the frames in the video.

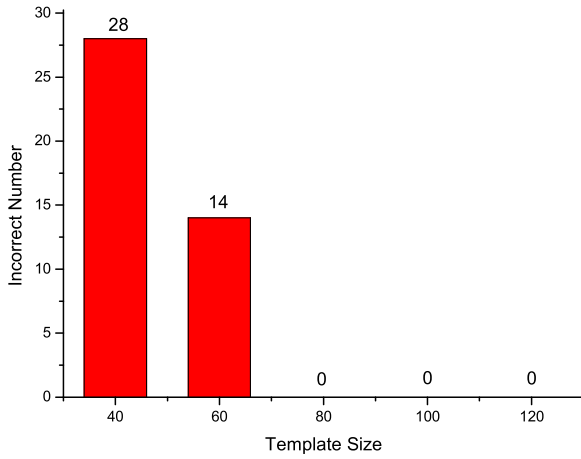


Figure 9. Numbers of incorrect cross-correlations with different template sizes.

the template size was 60×60 pixels, more retinal features with matching data were used to calculate the correlation, and the incorrect correlation number decreased to 14, but this still predicated a relatively inaccurate correlation. As the template size increased, the template included more features in the calculation of correlations, and it was clear that there were no incorrect correlations when the template size was 80×80 pixels.

Considering figures 7–9, a small template size led to an inaccurate cross-correlation, and the threshold value for template size was larger than 60×60 pixels in these experiments. A larger sized template contributed to cross-correlation precision as expected, but was associated with less computation efficiency. We summarized the correlation times for different template sizes in table 1. Note that the video had 877 frame sequences with a frequency of 160 Hz, which is much faster than a real-time record. In table 1, we can see that when the template size was smaller than 100×100 pixels and the calculation time of correlations was 30 Hz.

There are three main types of retinal motions: tremor, drift and microsaccades [33], and microsaccadic movements have a velocity of up to 300 deg s^{-1} , which is the hardest motion to be extracted due to their ultra-high frequency and large amplitudes. The template size (corresponding to the required minimum search area in the cross-correlation method) was determined by the spatial dimensions of the image pixels, and also the maximum anticipated retinal motion velocity and image frame rate [34]. In a LSO system retina imaging session, a square pixel spatial dimension of $14 \mu\text{m}$ per side of the retina, a maximum retinal motion velocity of 300 deg s^{-1} [35] and a frame rate of 160 Hz was achieved. Given that 1 deg is equivalent to approximately $291.5 \mu\text{m}$ on the human retina at the posterior pole [36], the template size in each direction in S pixels was defined as:

$$S = \frac{(300 \text{ deg s}^{-1}) \times (291.5 \mu\text{m deg}^{-1})}{(160 \text{ fps}) \times (14 \mu\text{m pixel})} \cong 40 \text{ pixels.} \quad (2)$$

Since S is the distance the retina can travel in any direction, the template window would have to be at least $2S$ to equal 80 pixels across. The maximum anticipated retinal motion velocity of 300 deg s^{-1} meant the template size was larger than 80×80 pixels. Considering experimental results plotted in figures 8–9, a template size larger than 60×60 pixels was needed to realize high accuracy cross-correlation calculations; on the other hand, the computational cost guarantee of cross-correlation operations required a template size smaller than 100×100 pixels. Therefore, a decisive conclusion for the template size was obtained: the template size was a compromise between enough image data on the retinal features to match reliably and having good temporal resolution. As an example result of this LSO system, the optimal template size was 80×80 pixels, which was suitable for a high speed and highly accurate cross-correlation to extract retinal motion.

4. Discussion and conclusions

This paper reported optimal targeting for retinal motion extraction using cross-correlation with a high speed LSO. Precision analysis of the cross-correlation matching experiments with LSO data was carried out to assess the performance of retinal motion extraction and to optimize the template target for the cross-correlation, including two key template parameters in the extraction of retinal motion; i.e., template location of retinal features and template size.

In this study, we found cross-correlation performance was highly sensitive to the template of retinal features, and the precision targeting of correlation templates were located on major retinal structures, such as vessel shadow patterns, because a vessel can be darker than the retinal background in normal fundus photographs, and the pattern of retinal vessels renders each frame unique, providing the most relevant information for the cross-correlation matching process. Particularly, the optimal target of a template was vessel bifurcation or vessel features occupying approximately eighty

Table 1. Comparison of different template sizes with cross-correlations^a.

Template size (in pixels)	40 × 40	60 × 60	80 × 80	100 × 100	120 × 120
Execute time	4.5 s	10.2 s	19.7 s	29.1 s	44.3 s

^a The cross-correlation was calculated using fast Fourier transform (FFT) methods and programmed with CUDA techniques of parallel processing, due to a GPU of NVIDIA GTX280 in connection with the host computer's CPU of E4600.

percent of the template area because the gray level difference between the vessel pattern and background was usually much larger than the difference between two non-corresponding areas in the retinal background.

In conjunction with determining that the optimal target for a cross-correlation template was retinal vessel bifurcation, the cross-correlation performance was quantified to determine the template size parameters. We found the optimized template size for our LSO system was 80 × 80 pixels and extracting retinal motions up to 300 deg s⁻¹ at speed of 30 Hz was possible.

Actually, there is no universal technique for solving precision targeting problems in a cross-correlation as it applies to extract retinal motion. The optimal template location and template size are highly dependent on the nature of the retinal images and system requirements. Nevertheless, the optimal template location and template size described here are appropriate and effective to extract retinal motion. In addition, the determination of cross-correlation templates can also be applied to other images having similar properties; i.e., relatively small features of distinct gray levels in an otherwise fairly uniform background.

Acknowledgments

We gratefully acknowledge the help of Professor Wenhan JIANG, Dr Jing LU and Dr Hao LI of the Institute of Optics and Electronics at the Chinese Academy of Sciences. This work is supported by the Outstanding Young Scientists of the Chinese Academy of Sciences, the National Science Foundation of China (Grant No. 61108082), the National Instrumentation Program (NIP, Grant No. 2012YQ120080), the National Hi-tech Support Project for the Twelfth Five Years' Blueprint (Grant No. 2012BAI08B04), the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KGCX2-Y11-920) and the West Light Foundation of the Chinese Academy of Sciences.

References

- [1] Riggs L A and Ratliff F 1952 The effects of counteracting the normal movements of the eyes *J. Opt. Soc. Am.* **42** 872–3
- [2] Ditchburn R W and Ginsborg B L 1952 Vision with a stabilized retinal image *Nature* **170** 36–7
- [3] Sperling G 1990 *Eye Movements and Their Role in Visual and Cognitive Processes* ed E Kowler (Amsterdam: Elsevier) pp 307–51
- [4] Martinez-Conde S, Macknik S L and Hubel D H 2004 The role of fixational eye movements in visual perception *Nat. Rev. Neurosci.* **5** 229–40
- [5] Roorda A and Williams D R 1999 The arrangement of the three cone classes in the living human eye *Nature* **397** 520–2
- [6] Fiolka R 2013 Three-dimensional live microscopy beyond the diffraction limit *J. Opt.* **15** 094002
- [7] Li S, Kuang C, Hao X, Wang Y, Ge J and Liu X 2013 Enhancing the performance of fluorescence emission difference microscopy using beam modulation *J. Opt.* **15** 125708
- [8] Huang D *et al* 1991 Optical coherence tomography *Science* **254** 1178–81
- [9] Shevchuk A I, Novak P, Velazquez M A, Fleming T P and Korchev Y E 2013 Combined ion conductance and fluorescence confocal microscopy for biological cell membrane transport studies *J. Opt.* **15** 094005
- [10] Chen S-Y, John Hsu Y, Yeh C-H, Chen S-W and Chuang C-H 2015 Pico-projector based optical sectioning microscopy for 3D chlorophyll fluorescence imaging of mesophyll cells *J. Opt.* **17** 035301
- [11] Li H, Lu J, Shi G H and Zhang Y D 2010 Tracking features in retinal images of adaptive optics confocal scanning laser ophthalmoscope using KLT-SIFT algorithm *Biomed. Opt. Express* **1** 31–40
- [12] Akagi-Kurashige Y *et al* 2014 Retinal microstructural changes in eyes with resolved branch retinal vein occlusion: an adaptive optics scanning laser ophthalmoscope study *Am. J. Ophthalmol.* **157** 1239–49
- [13] Zhang Y, Wang X, Rivero E B, Clark M E, Witherspoon C D, Spaide R F, Girkin C A, Owsley C and Curcio C A 2014 Photoreceptor perturbation around subretinal drusenoid deposits as revealed by adaptive optics scanning laser ophthalmoscopy *Am. J. Ophthalmol.* **158** 584–96
- [14] Yang Q, Arathorn D W, Tiruveedhula P, Vogel C R and Roorda A 2010 Design of an integrated hardware interface for AOSLO image capture and cone-targeted stimulus delivery *Opt. Express* **18** 17841–58
- [15] Deák G G and Schmidt-Erfurth U 2013 Imaging of the parafoveal capillary network in diabetes *Curr. Diab. Rep.* **13** 469–75
- [16] Park S P, Chung J K, Greenstein V, Tsang S H and Chang S 2013 A study of factors affecting the human cone photoreceptor density measured by adaptive optics scanning laser ophthalmoscope *Exp. Eye Res.* **108** 1–9
- [17] Cornsweet T N 1958 New technique for the measurement of small eye movements *J. Opt. Soc. Am.* **48** 808–11
- [18] O'Connor N J *et al* 1998 Fluorescent infrared scanning-laser ophthalmoscope for three-dimensional visualization: automatic random-eye-motion correction and deconvolution *Appl. Opt.* **37** 2021–33
- [19] Stevenson S B and Roorda A 2005 Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy *Proc. SPIE* **5688** 145–51
- [20] Vogel C, Arathorn D, Roorda A and Parker A 2006 Retinal motion estimation in adaptive optics scanning laser ophthalmoscope *Opt. Express* **14** 487–97
- [21] Sato Y, Chen J, Zoroofi R A, Harada N, Tamura S and Shiga T 1997 Automatic extraction and measurement of leukocyte

- motion in microvessels using spatiotemporal image analysis *IEEE Trans. Biomed. Eng.* **44** 225–36
- [22] Kraus M F, Potsaid B, Mayer M A, Bock R, Baumann B, Liu J J, Hornegger J and Fujimoto J G 2012 Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns *Biomed. Opt. Express* **3** 1182–99
- [23] He Y, Li H, Lu J, Shi G H and Zhang Y D 2013 Retina imaging by using compact line scanning quasiconfocal ophthalmoscope *Chin. Opt. Lett.* **11** 021101
- [24] He Y, Shi G H, Lu J, Li H and Zhang Y D 2013 High-speed line scanning confocal laser ophthalmoscope *Acta Opt. Sin.* **32** 0117001
- [25] He Y, Li H, Lu J, Shi G H and Zhang Y D 2011 Optical design of a line scanning laser ophthalmoscope *Am. J. Eng. Technol. Res.* **11** 812–6
- [26] Eli P, Augliere R A and Timberlake G T 1987 Feature-based registration of retinal images *IEEE Trans. Med. Imaging* **MI-6** 272–8
- [27] Penney G P et al 1998 A comparison of similarity measures for use in 2D–3D medical image registration *IEEE Trans. Med. Imaging* **17** 586–95
- [28] Sheehy C K, Yang Q, Arathorn D W, Tiruveedhula P, de Boer J F and Roorda A 2012 High-speed, image-based eye tracking with a scanning laser ophthalmoscope *Biomed. Opt. Express* **3** 2611–22
- [29] Nagin P, Schwartz B and Raynolds G 1985 Measurement of fluorescein in angiograms of the optic disc and retina using computerized image analysis *Ophthalmology* **92** 547–52
- [30] He Y, Wang Z, Wei L, Li X, Shi G and Zhang Y 2014 Compact high-speed line scanning quasi-confocal ophthalmoscope and retina imaging experiments *Proc. SPIE* **9230** 923024
- [31] Barrett S F, Jerath M R, Rylander H G III and Ashley J W 1994 Digital tracking and control of retinal images *Opt. Eng.* **33** 150–9
- [32] He Y, Wei L, Wang Z, Yang J, Li X, Shi G and Zhang Y 2015 Extraction of ultra-high frequency retinal motions with a line scanning quasi-confocal ophthalmoscope *J. Opt.* **17** 015301
- [33] Spauschus A, Marsden J, Halliday D M, Rosenberg J R and Brown P 1999 The origin of ocular microtremor in man *Exp. Brain. Res.* **126** 556–62
- [34] Wright C H G, Ferguson R D, Rylander H G III, Welch A J and Barrett S F 1997 Hybrid approach to retinal tracking and laser aiming for photocoagulation *J. Biomed. Opt.* **2** 195–203
- [35] Martinez-Conde S, Macknik S L, Troncoso X G and Hubel D H 2009 Microsaccades: a neurophysiological analysis *Trends Neurosci.* **32** 463–75
- [36] Vaughan D G, Asbury T and Riordan-Eva P 1992 *General Ophthalmology* 13th edn (Norwalk, CT: Appleton and Lange)