IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2019, accepted April 23, 2019, date of publication May 3, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914463

Automatic Dewarping of Retina Images in Adaptive Optics Confocal Scanning Laser Ophthalmoscope

HAO CHEN[1,2], YI HE[2], LING WEI[2], XIQI LI[2], AND YUDONG ZHANG[2]

[1]College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China
[2]Key Laboratory on Adaptive Optics, Institute of Optics and Electronics, Chengdu 610209, China

Corresponding authors: Yi He (heyi_job@126.com) and Ling Wei (frank_starcraft@163.com)

This work was supported in part by the National Science Foundation of China under Grant 61605210, in part by the National Key Research and Development Program of China under Grant 2016YFC0102500 and Grant 2017YFB0403700, and in part by the National Instrumentation Program (NIP) under Grant 2012YQ120080.

ABSTRACT Retina images acquired by an adaptive optics confocal scanning laser ophthalmoscope (AOSLO) usually need to remove image warp to improve image quality. The most significant task of AOSLO image dewarping is image registration. Most traditional feature-based registration algorithms used for AOSLO images are based on Gaussian scale space. However, the homogeneous Gaussian blurring reduces the localization accuracy of feature points and the distinctiveness of feature descriptors. In this paper, the accelerated KAZE (AKAZE) feature based on nonlinear scale space was utilized to register AOSLO retinal images for the first time, and an efficient strategy based on matched feature points for frame selection was proposed to automatically accomplish AOSLO retinal image dewarping. Moreover, a flexible method based on power spectra analysis is proposed to study the minimum number of frames needed to accomplish image dewarping. The extensive experiments demonstrated that the AKAZE method is more suitable for AOSLO image dewarping, benefitted with better accuracy, robustness, and rapidity compared with several traditional registration methods based on Gaussian scale space.

INDEX TERMS Retina image, adaptive optics confocal scanning laser ophthalmoscope, image dewarping, image registration, nonlinear scale space, Gaussian scale space.

I. INTRODUCTION

The adaptive optics (AO) technique can efficiently compensate for ocular aberrations in the eye's optics [1], and therefore it is helpful to obtain retina images with near diffraction-limited resolution. Microscopic imaging of the living human retina at the single-cell level has been achieved with AO in the past decades [2]–[9]. AO has also been combined with the confocal scanning laser ophthalmoscope (AOSLO) to obtain high-resolution retina images [3], [6], [7], [10]. However, the high resolution of AO retina-imaging systems has been affected by fixational eye movements, which includes various components from a low to relatively high frequency of 150 Hz [11], [12]. Therefore, even though the object you are looking at seems to be fixed in your field of view, the image of the object sweeps across dozens of photoreceptors within milliseconds. Since the AOSLO system usually works with small field of views (approximately 0.75 to 1.5 degrees), the distortion is dominated by eye movements. To improve the signal-to-noise ratio (SNR) of the image, we usually need to average multi-frame AOSLO images together. However, due to the eye's movement, we will only get a motion-induced blurry image. Thus, eliminating the effect of retina motion is an important task to improve the quality of retina images. The most significant part of the task is achieving accurate image alignment before superposing them together.

There are several registration algorithms that have been applied to AOSLO image processing. For example, Stevenson and Roorda [13] used a patch-based cross-correlation method for retina motion detection, which could only estimate the drift. Vogel *et al.* [14] used the map-seeking circuit algorithm to estimate the motion of the eye's retina. Faisan *et al.* [15] adopted a cross-correlation method to find matched points, and used a multi-scale B-spline representation to map images to each other accurately; however, it was very

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

time-consuming. All these methods based on cross-correlation matching have some distinct disadvantages. The other kind of registration method applied to AOSLO image processing is the feature-based method. Chen *et al.* [16] proposed a novel highly distinctive local feature descriptor named the partial intensity invariant feature descriptor (PIIFD) to accomplish multi-modal retina image registration. However, it fails to register retina images with other modalities in the presence of large content or scale changes. The scale invariant feature transform (SIFT) method was firstly adopted for AOSLO image registration by Li *et al.* [17], [18]. Then, Ghassabi *et al.* [19] combined the SIFT and PIIFD together to achieve more accurate registration. Li *et al.* [20] proposed the PCA-SIFT method to achieve automatic AOSLO image stitching. Moreover, Chen *et al.* [21] used the SIFT feature to achieve automatic retina image stitching with multi-modal images. Although the SIFT feature is relatively accurate, its real-time performance is poor. To accelerate the SIFT algorithm, He *et al.* [22] used the fast Fourier transform to build a Gaussian pyramid, and their improved SIFT method was 2–3 times faster than the original one. In addition, the speeded-up robust features (SURF) has also been used for AOSLO images by Li *et al.* [23]. Although the SURF algorithm is much faster than SIFT, it pays the price of a loss in performance. Furthermore, Davidson *et al.* [24] utilized oriented FAST, rotated BRIEF (ORB) method to accomplish fast multi-modal retina images montaging. Recently, Cao *et al.* [25] improved the SURF method with geometric algebra (GA-SURF) and applied it to medical images. All of these approaches and the many related algorithms that have followed (e.g. the binary robust invariant scalable keypoints (BRISK) [26]) rely on the use of Gaussian scale space (GSS) and sets of Gaussian derivatives as smoothing kernels for scale space analysis. However, a serious problem is that GSS does not respect the natural boundaries of objects and smoothes to the same degree both details and noise at all scale levels, which finally results in a loss of localization accuracy of feature points and a decrease in the distinctiveness of feature descriptors [27].

To overcome these problems, some methods have been recently presented. They aim to detect and describe features in nonlinear scale space (NSS) [28]–[30]. The main idea of these methods is making blurring locally adaptive to the image data, blurring small details but preserving object boundaries. Therefore, the locations of the features are more accurate and the corresponding feature descriptors are more reliable. Wang *et al.* [30] used the bilateral filter to build the NSS with a higher computational complexity, making it unsuitable for a real-time image registration application. The KAZE feature was introduced by Alcantarilla *et al.* [29]. The main drawback of KAZE is the high computation complexity in building NSS. Recently, Alcantarilla and Solutions [28] proposed a so-called accelerated-KAZE feature (AKAZE). The main contributions of their paper were that they designed a much more efficient scheme to build the NSS and improved the feature descriptor by introducing a binary descriptor.

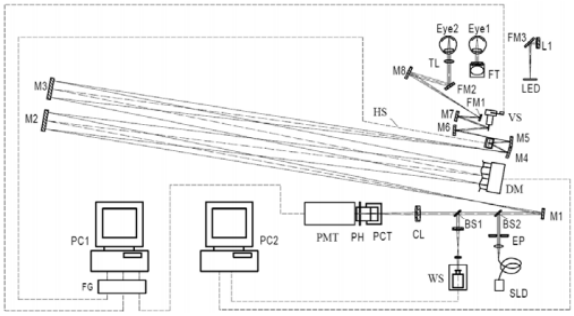


FIGURE 1. The schematic of AOSLO system. VS, vertical scanner; WS, wavefront sensor; DM, deformable mirror; PMT, photomultiplier tube; BS, beam splitter; CL, collecting lens; FM, folding mirrors; SLD, superluminescent laser diode; M1∼M8, spherical mirrors; PH, pinhole; TL, trial lens; FG, Frame Grabber; LED, LED array; L1, magnifying lens; PC1, computer for image acquisition; PC2, computer for adaptive optics control; FT, fixing target.

More recently, with the development of deep learning technique, some registration methods based on convolutional neural network (CNN) have also been proposed [31]–[33]. The largest advantage of CNN-based methods is that it is easily to automatically obtain the transformation parameters once the CNNs have been well trained. As we know, the CNN-based registration methods are data-driven, therefore it is heavily relied on the quality and quantity of training datasets. To ensure the convergence of CNNs, the training datasets should have relatively good signal-to-noise ratio. On the other hand, large amount of datasets are needed to ensure the robustness of the trained CNN, which is not suitable for such an experimental system.

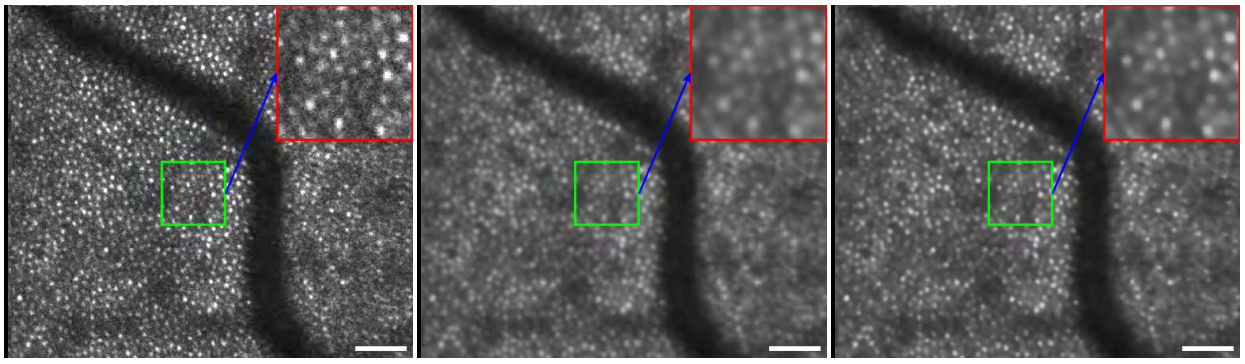
In this paper, the accelerated KAZE (AKAZE) feature based on NSS was utilized to register AOSLO retinal images for the first time, and an efficient strategy based on matched feature points for frame selection was proposed to automatically accomplish AOSLO retinal image dewarping. Moreover, a flexible method based on power spectra analysis is proposed to study the minimum number of frames needed to improve image quality.

The remainder of the paper is organized as follows. In Section II, we introduce our AOSLO imaging system, and then we compare the AOSLO images blurred by Gaussian kernels and nonlinear diffusion filters. In Section III, we describe the AOSLO image registration procedure of the AKAZE algorithm. Afterwards, we perform an extensive analysis of different registration algorithms' performances on both synthetic and real AOSLO images data in Section IV. In addition, we study both the frame selection strategy and the minimum number of frames needed to improve image quality. In Section V we discuss the merits and limitations of the AKAZE algorithm. Finally, we conclude this paper with closing observations and a discussion of future directions in Section VI.

II. AOSLO IMAGING AND IMAGE BLURRING

A. AOSLO SYSTEM FOR RETINA IMAGING

An AOSLO system for real-time (30Hz) retina imaging was set up in our laboratory [7]. Fig. 1 shows a schematic diagram



(a) (b) (c)

FIGURE 2. A typical AOSLO image blurred by different filters. (a) A retina image acquired by the AOSLO system with 1ř FOV in a 30-year-old healthy male; (b) blurring (a) with a Gaussian filter; (c) blurring (a) with a nonlinear diffusion filter. The sub-images in green box were enlarged to red box with scale equaling 2. Blurring scale $\sigma = 6.4$. The scale bar is 50 μm.

of the setup. The light beam emitted from a superluminescent laser diode (SLD, with a central wavelength of 795nm, SLD-381-HP, supernum, Ireland) is collimated, and then is focused to a small spot on retina by optic system of human eye. Light scattered back from retina is split into two parts. One part passing through a confocal pinhole placed optically conjugate to the retina is collected by a photomultiplier tube (PMT, H7422-20, Hamamastu, Japan). And the other part is captured by a Shack-Hartmann wavefront sensor. The slope data of the wavefront is collected and processed by the computer. The aberration correction is performed using a 61-actuator deformable mirror (DM) with an aperture of 20 mm. Two scanning mirrors (horizontal scanner: 16 KHz resonant scanner, vertical scanner: 30Hz galvometric scanner) are used to control the focused spot to scan the retina. By synchronizing the PMT signal and two scanning mirrors, the image of retina can be consecutively recorded.

With the help of AOSLO, we could obtain nearly diffraction-limited high resolution retina images, as presented in Fig. 2 (a) (30 years old, healthy male). The frame size is 512×449 pixels, the intensity of the frame is between [0,255] and the field of view (FOV) is 1 degree (about 398 μm). Each frame has a pixel resolution of 0.78 μm/pixel assuming an axial length of 22.8 mm.

B. BLURRING AOSLO IMAGES WITH DIFFERENT FILTERS

To improve the image quality of AOSLO images, we usually need to align consecutively captured AOSLO images and average them together. Traditional feature-based registration methods applied to AOSLO images processing are based on GSS. However, Gaussian filtering does not respect the natural boundaries of objects, and smoothes to the same degree both details and noise at all scale levels. To overcome these problems, methods based on NSS have been proposed. The main idea of these methods is making blurring locally adaptive to the image data, blurring small details but preserving object boundaries. One of the most representative nonlinear diffusion filtering equations is the well-known Perona-Malik equations [27], and it is defined as

$$\begin{cases} \dfrac{\partial u}{\partial t} = \nabla \left(g \left(|\nabla u|^2 \right) \nabla u \right) \\ u\left(x, y, 0\right) = u_0\left(x, y\right). \end{cases} \tag{1}$$

Here, $\partial u/\partial t$ denotes the blurred image at evolution time t, t is directly related to the blurring scale, x and y denote the horizontal and vertical coordinates of the image, respectively; and u_0 is the original image. $g\left(|\nabla u|^2\right) \in [0, 1]$ is the diffusivity coefficient function which is also termed as the edge-stopping function (ESF) [34], and it is a monotonically decreasing function of the magnitude of gradient.

Fig. 2 shows the blurred images at the same scale with different filters. As seen in Fig. 2, it is obvious that the nonlinear diffusion filtering could preserve more image features, such as the edges of vessels and the structures of cones whereas Gaussian filtering equally smoothes all the structures in the image. To evaluate the blurring effect, the normalized residual (NR) values is utilized. NR can be calculated by,

$$NR = \frac{\|B - O\|_2}{N}. \tag{2}$$

Here, B denotes the blurred image, O is the original image, and N is the number of pixels in O. The smaller the NR value, the closer the blurred image is to the original image. Table 1 presents the NR values of blurred images obtained by filtering Fig. 2 (a) with different filters at different scales.

In Fig. 2 and Table 1, it is evident that nonlinear diffusion filtering can preserve more features and their nearby information for AOSLO images. Therefore, we believe that features extracted in NSS ought to be more accurate, repeatable, and distinctive.

III. AKAZE FEATURES

As AKAZE method is one of the most representative registration methods based on the NSS, we briefly introduce it in this section. Similar to traditional methods, AKAZE algorithm mainly includes 3 parts: building image scale space, detecting feature points and constructing feature descriptors [30].

TABLE 1. The normalized residual values of different blurred images. The values of second column correspond to Fig. II (b) and (c) respectively.

Blurring Scale σ [a]	3.2	6.4	12.8
Gaussian filtering	0.0752	0.0824	0.0894
Nonlinear diffusion filtering	0.0709	0.0742	0.0781

[a]σ is the blurring scale. For the Gaussian filter, the blurring scale is the standard deviation of the filter; for the nonlinear diffusion filter, the blurring scale σ is directly related to the evolution time t (as given in Eq. (1)), the relationship is $t = \sigma^2/2$.

However, the biggest difference between AKAZE method and other methods is that it can efficiently build the NSS with a mathematical framework called Fast Explicit Diffusion (FED) [35], [36].

By means of the FED scheme, a NSS can be built much faster than with any other kind of discretization scheme. Once the last sublevel in each octave is calculated, the filtered image is downsampled with a factor of 2 using the pixel area relation, and the downsampled image is used as the starting image for the next FED cycle in the next octave. Then, it is straightforward to build the NSS with the FED scheme.

Once the NSS is established, feature points can be efficiently detected. Different from SIFT algorithm, AKAZE method computes the response of a scale-normalized determinant of the Hessian at multiple scale levels,

$$Det_{Hessian} = \sigma^2 \left(D_{xx} D_{yy} - D_{xy}^2 \right). \tag{3}$$

where D_{xx} and D_{yy} are the second-order horizontal and vertical derivatives respectively, D_{xy} is the second-order cross derivative, and σ denotes the blurring scale introduced in the footnote of Table 1. Given the set of filtered images from the NSS, the detector responses at different scale levels are obtained. Then the maxima are searched in scale and spatial locations. Finally, as proposed by Lowe [37] and Bay *et al*. [38], the position of the keypoint is refined with sub-pixel accuracy.

For feature descriptors construction, the AKAZE method proposed a modified-LDB [39] (MLDB) scheme by subsampling the grids in steps that are a function of the scale σ of the feature.

IV. EXPERIMENTS

Ten eyes of eight normal volunteers were included in this study, and ten datasets were acquired for the study. All light exposures adhered to the maximum permissible exposure limits set by the American National Standards Institute standards [40]. Subjects' eyes were dilated with phenylephrine hydrochloride (2.5%) and tropicamide (1%). Subjects were instructed to fixate (using the imaged eye) as steadily as possible at a target while AOSLO videos were acquired. However, since the retina is always in motion, the imaging system cannot stably observe the region of interest for a long time. For our experimental system, the stable observation time averagely was not smaller than 4 seconds, and therefore at least 100 stable frames were extracted from each dataset, then it means that over 1000 frames were utilized to validate the performances of different registration methods.

As mentioned in Section I, AOSLO images suffer from retina motion. In addition, the SNR of AOSLO image is very low due to various sources of noise and a safer laser illumination. To improve the image quality, eliminating the motion and noise is necessarily. One classical way to achieve it is superposing consecutive multi-frame AOSLO images after registration [13]–[15], [17], [18], [41]–[43].

In this section, four registration algorithms (SIFT, SURF, ORB, and AKAZE) are used to accomplish AOSLO image dewarping. To evaluate the performances of these methods, two popular metrics, the precision value (PR) and matching score (MS) introduced in [44], [45], are utilized. The PR value is computed by the following equation:

$$PR = \frac{\#correct\ matches}{\#total\ tentative\ matches}. \tag{4}$$

The PR value means how relevant the matched features are to each other, and a higher PR value represents better distinctiveness among matched features.

The MS value is obtained with the following formula:

$$MS = \frac{\#correct\ matches}{\#Features}. \tag{5}$$

The MS value describes the correct matches in the initial features, and the highest MS value indicates the method with the best performance. It also reflects the sensitivity of the investigated methods.

The above metrics mainly concentrate on evaluations of the distinctiveness and repeatability of different feature detection methods. To evaluate the alignment accuracy of different methods, we should use the information of overlapping regions. The normalized cross-correlation (NCC) measure is utilized in this paper, which is widely used as a measure of image alignment [46], [47]. The NCC value can be calculated by

$$NCC = \frac{1}{N} \frac{\sum_{x,y \in \Psi} \left(R(x,y) - M_R \right) \left(Q(x,y) - M_Q \right)}{\sigma_R \times \sigma_Q}. \tag{6}$$

Here, R and Q denote the reference image and the query image, respectively. Ψ is the overlapping area of the two images. M_R and M_Q represent the mean intensities of the overlapping area of R and Q, respectively. N is the total number of pixels in the overlapping area. σ_R and σ_Q denote the standard deviations of pixels in the overlapping area of R and Q, respectively. The NCC value ranges from -1 to 1, where 1 represents perfect alignment. The second evaluation measure is the normalized mutual information (NMI) metric [48], given by

$$NMI = \frac{H(R) + H(Q) - H(R,Q)}{\sqrt{H(R) H(Q)}}. \tag{7}$$

where

$$H(R) = -\sum_{i}^{W} P_R(R_i) \log P_R(R_i). \quad (8)$$

$$H(Q) = -\sum_{j}^{V} P_Q(Q_j) \log P_Q(Q_j). \quad (9)$$

and

$$H(R, Q) = -\sum_{i}^{W}\sum_{j}^{V} P_{R,Q}(R_i, Q_j) \log P_{R,Q}(R_i, Q_j). \quad (10)$$

are the marginal (Eq. (8) and Eq. (9)) and the joint (Eq. (10)) entropy measures. $P_R(R_i)$ and $P_Q(Q_j)$ describe the marginal probability density estimate of the W and V possible image intensities R_i and Q_j in image R and image Q, respectively. $P_{R,Q}(R_i, Q_j)$ denotes the joint probability density estimate correspondingly. The measure ranges from 0 to 1, with 1 representing perfect alignment. NMI is a statistical evaluation of the image intensities in each image and their joint relationship. One primary advantage of NMI is that, unlike the NCC, it does not assume a linear relationship between the image intensities in the two images.

Both NCC and NMI are intensity-based similarity metrics. Although intensity-based similarity metrics are influenced by high correlations between neighboring pixels [49], the intensity-based similarity metrics have been proved successful for intra-modal case where images are similar on intensity and texture, as pointed out in the relevant literatures [50], [51]. As the AOSLO images are consecutively acquired, it means that the average intensities of these images are similar or vary in a small range. In addition, there are similar textures in all frames due to the special structures of cone photoreceptors (honeycomb structure). Therefore, the NCC and NMI metrics can well reflect the alignment accuracy and have been widely-used for AOSLO images similarity evaluation [21], [24].

Furthermore, to provide a more complete evaluation, a human vision system based (HVS-based) similarity index termed as structural similarity index (SSIM) is utilized [52]. The SSIM measure is defined as,

$$SSIM(R, Q) = \frac{(2M_R M_Q + C_1)(2\sigma_{RQ} + C_2)}{(M_R^2 + M_Q^2 + C_1)(\sigma_R^2 + \sigma_Q^2 + C_2)} \quad (11)$$

where C_1 and C_2 are two small positive constants, σ_{RQ} is the cross deviation. The maximum SSIM index value 1 is achieved if and only if R and Q are identical. Different from intensity-based metrics, SSIM provides good image quality prediction performance for a wide variety of image distortions [52], [53].

Finally, since we have no knowledge about the true warpless and noise-less image, we use the power spectrum and image contrast to evaluate the quality of the superposed image. The image contrast [54] is defined as

$$Contrast = \frac{\sigma_I}{\bar{I}}. \quad (12)$$

where σ_I denotes the standard deviation of intensities of the image I, and $\bar{I}$ is the mean intensity of the image.

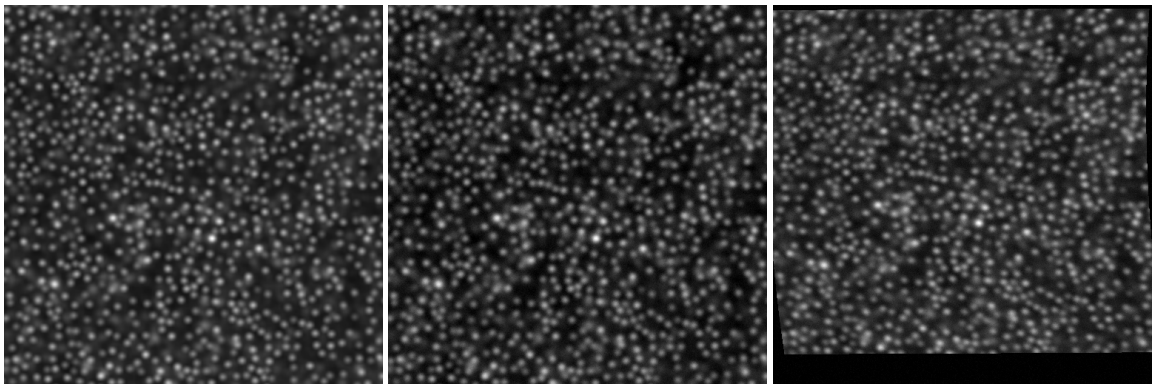
The following experiments and corresponding results were accomplished on an Intel Core i5-2430 CPU 2.4GHz laptop computer with OpenCV implementations of SIFT, SURF, ORB, and AKAZE.

A. SIMULATION VERIFICATION OF DIFFERENT REGISTRATION METHODS ON SYNTHETIC AOSLO IMAGES

As mentioned above, the retina is always in movement even when you have fixation on a target, and retina motions include various components from a low to relatively high frequency of 150 Hz. Although the relatively high imaging rate of AOSLO (30 Hz or even faster in other systems [55], [56]) makes it possible to approximately align retina images, the real relationship between any two frames is impossible to obtain due to the small intra-frame distortions. Furthermore, anatomical features which are independently annotated across images cannot be found or be precisely located in AOSLO images. For example, the typical structure of cone photoreceptor is a small saturated block in the image, and it is hard to find the center of the cone with pixel or sub-pixel accuracy. In addition, there are small differences between corresponding cones in two images due to the retina motions, aberration changes and various sources of noise. These facts imply that even approximately true transformations cannot be obtained through the calibration-based method.

To achieve a reliability validation, we propose to generate synthetic AOSLO images and carry out experiments on these data. Firstly, a set of ideal retina images are created and normalized by using the algorithm described in reference [57]. Secondly, for each ideal image, a set of point spread functions (PSF) are generated to simulate the residual optical aberration of the eye, as presented in reference [58]. Then, the blurred image is generated by convolving with the PSFs. Once the blurred image is obtained, a set of second-order polynomial transformation matrixes are created according to the statistical results presented in reference [17], then a simulated dataset is obtained by multiplying the transformation matrixes and the blurred image. Finally, Gaussian noise with standard deviation 0.015 is also added to the dataset [58].

Fig. 3 shows some of the synthetic images. For each dataset, we take the first frame as the reference image, and the registration algorithms are applied to the reference image and remaining query images. The total tentative matched feature points are determined by the nearest neighbor distance ratio matching criteria [37] with a distance ratio value equaling 0.8. Using the ground-truth transformation, a tentative match is considered as a correct match when the error (the distance between the transformed position and detected position) is less than 2.5 pixels (corresponding to the radius of the minimum cone photoreceptor). One thing to be clarified here is that since the performances of these algorithms are sensitive to the number of keypoints in every image, to be



(a) (b) (c)

FIGURE 3. Synthetic AOSLO images. (a) Original image (eccentricity 1.5 mm from foveal center); (b) degraded image of (a); (c) one transformed image of (b).

TABLE 2. The averaged values (Ave) of Precision Value (PR) and Matching Score (MS) on synthetic datasets in different positions of retina. The largest value in each column is highlighted in bold.

Eccentricity	0.5 mm		1.0 mm		1.5mm	
Ave %	PR	MS	PR	MS	PR	MS
SIFT	99.13	62.09	99.40	**63.34**	98.86	59.66
SURF	97.68	56.37	97.13	60.14	97.50	56.51
ORB	95.37	49.84	93.74	44.33	93.34	45.34
AKAZE	**99.14**	**63.66**	**99.59**	63.23	**99.40**	**60.80**

as fair as possible, all these and all following results are obtained with the average number of keypoints in every image equaling 1000.

The averaged values of PR and MS metrics are exhibited in Table 2, and it is obvious that the AKAZE algorithm obtains the best PR and MS values in all cases, except for the synthetic AOSLO images with an eccentricity of 1.0 mm, in which the AKAZE method produces a better PR value but a smaller MS value than the SIFT method. Actually, the MS values for these two methods are quite close and no significant difference can be achieved. With the increase of eccentricity, the performance of the AKAZE method becomes more superior compared to other three methods.

Before accomplishing the registration process on synthetic data, an important consideration is highlighted here. In order to simulate the registration process for real AOSLO images, the random sample consensus (RANSAC) [59] algorithm is utilized. Truly RANSAC cannot make sure to remove all outliers and preserve all inliers in theory. In addition, the reliability of RANSAC with affine transformation model is confusing to select correct matches for second-order polynomial transformation model. But RANSAC indeed works well in AOSLO image dewarping for two reasons. The first reason is that the frames are continually acquired, and the subjects are instructed to fixate (using the imaged eye) as steadily as possible at a target, then the deformation between two frames are relatively small (as presented in Table 1 in reference [17], the parameters of the second-order term are very small). It means that outliers removed by RANSAC with affine transformation model is also the outliers in second-order polynomial transformation model with very high probability; Secondly, once a suitable threshold (e.g. equaling 10, as suggested by Davidson *et al.* in [24]) is used in the RANSAC method, then it is efficiently to remove almost all outliers while preserve most of inliers. Table 3 presents the change of inliers and outliers after running RANSAC method, and it is obvious that RANSAC method can remove almost all outliers and preserve most of inliers (99.8%) for SIFT, SURF and AKAZE methods. For ORB method, the performance of RANSAC is relatively worse. The results indicate that RANSAC with affine transformation model also works well for second-order polynomial transformation model in the case of AOSLO image dewarping.

We also provide the average values and standard deviations of NCC, NMI and SSIM values in Table 4. In Table 4, it is obvious that the AKAZE method achieves better or competitive performance compared to the SIFT method, and both of them have better performance compared to other two methods. It seems that the differences between these methods are relative small, but these results are obtained at an ideal situation.

B. PERFORMANCES OF DIFFERENT REGISTRATION METHODS ON REAL AOSLO IMAGES

Different from the synthetic data, the real AOSLO images have more complex structures and suffer from various sources of noise. In addition, the difference of residual aberrations between the reference image and query image becomes large with the query frame departing from the reference frame, making it hard to achieve accurate alignment. Therefore, it is important to evaluate the performance of different registration methods on real AOSLO image dewarping. Since at least six corresponding point pairs are needed to implement the second-order polynomial transformation, if there

TABLE 3. The changes of inliers and outliers and their corresponding occupancy ratios with RANSAC method.

	Inlier matches before RANSAC		Inlier matches after RANSAC	
Ave	Number	Occupancy Ratio (%)	Number	Occupancy Ratio (%)
SIFT	583	98.85	566	99.79
SURF	506	96.59	494	99.89
ORB	412	92.96	401	95.18
AKAZE	582	99.18	567	99.83
	Outlier matches before RANSAC		Outlier matches after RANSAC	
Ave	Number	Occupancy Ratio (%)	Number	Occupancy Ratio (%)
SIFT	6.65	1.15	1.07	0.21
SURF	17.36	3.41	0.50	0.11
ORB	30.40	7.04	19.93	4.82
AKAZE	4.65	0.82	0.91	0.17

TABLE 4. The averaged values (Ave) and standard deviations (Std) of normalized cross-correlation (NCC), normalized mutual information (NMI) and structure similarity index (SSIM) values on synthetic datasets in different positions of retina. The largest value of Ave and smallest value of Std in each column are highlighted in bold.

Eccentricity	0.5 mm			1.0 mm			1.5mm		
Ave (Std) %	NCC	NMI	SSIM	NCC	NMI	SSIM	NCC	NMI	SSIM
SIFT	98.04/0.33	37.48/0.62	93.10/0.47	98.48/**0.33**	39.49/**0.40**	92.45/0.44	98.30/0.28	37.40/0.19	89.83/0.46
SURF	98.04/0.33	37.45/0.57	93.10/0.46	98.48/0.33	39.48/0.42	92.45/**0.42**	98.30/0.28	37.35/0.37	89.82/0.47
ORB	97.91/0.42	36.40/1.00	92.84/0.47	98.29/0.41	38.09/0.96	92.20/0.44	98.10/0.42	36.12/0.96	89.60/0.46
AKAZE	**98.07**/**0.32**	**37.49**/**0.40**	**93.11**/**0.43**	**98.49**/0.34	**39.50**/0.41	**92.45**/0.45	**98.33**/**0.28**	**37.42**/**0.19**	**89.83**/**0.46**

TABLE 5. The mean values (Mean) and standard deviations (Std) of Precision Value (PR) and Matching Score (MS) on ten datasets. The largest values of columns are highlighted in bold.

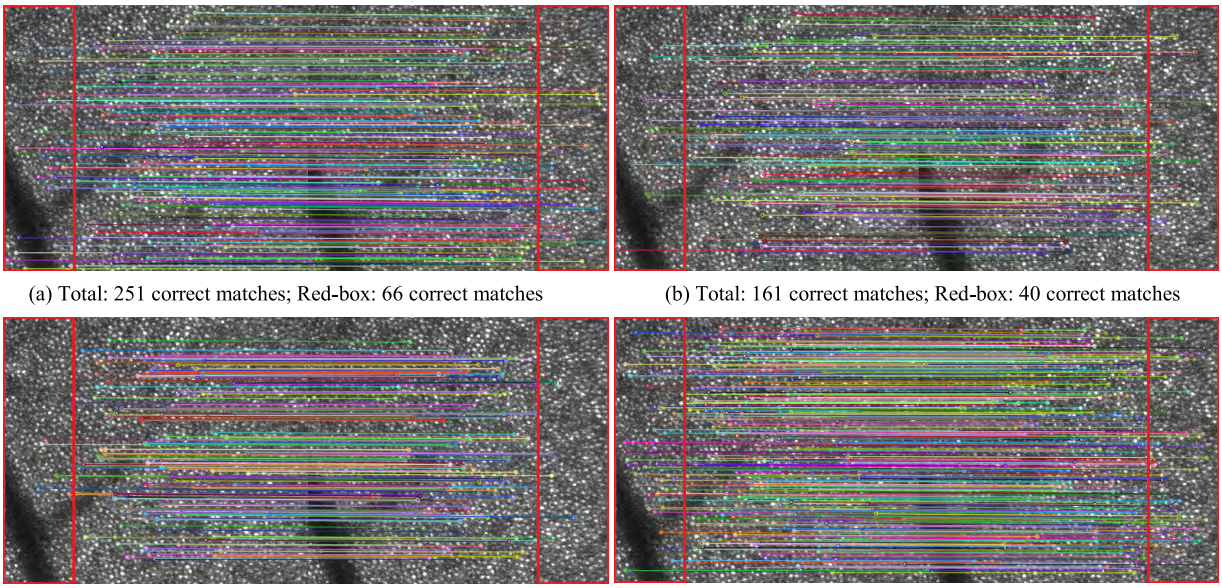
Precision Value											
Mean/Std %	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	67/**22.5**	70/19.4	76/15.4	78/15.6	83/11.8	80/**12.5**	78/12.2	72/**14.8**	72/**17.9**	74/16.9	74.9/15.9
SURF	54/17.4	51/**20.6**	56/**17.6**	54/**21.6**	59/**17.4**	68/11.9	63/**13.2**	53/12.7	62/16.8	59/**18.5**	57.8/**16.8**
ORB	94/5.1	**93**/4.9	**95**/5.2	**95**/4.6	93/5.3	**95**/3.6	**95**/4.2	**94**/4.4	**94**/4.9	**95**/4.5	**94.3**/4.7
AKAZE	92/6.5	**93**/4.4	94/5.6	94/5.4	**94**/4.7	94/5.2	93/6.2	**94**/4.8	93/5.2	93/6.5	93.3/5.4
Matching Score											
Mean/Std %	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	7.1/6.0	6.4/5.8	8.1/6.7	8.7/6.8	10.7/7.5	10.2/6.2	8.8/6.6	5.9/4.4	7.4/5.8	7.7/6.1	8.1/6.2
SURF	6.4/4.2	5.5/4.3	6.3/4.3	6.1/4.6	7.0/4.6	8.8/4.3	7.8/4.7	5.1/3.0	7.6/4.7	6.9/4.7	6.8/4.3
ORB	**21.7**/**8.7**	16.9/7.1	17.4/5.6	17.1/5.5	14.6/5.5	23.6/6.9	22.2/6.1	15.3/5.7	**24.7**/6.8	23.4/7.3	19.7/6.5
AKAZE	19.8/8.2	**18.6**/**8.4**	**21.2**/**7.3**	**21.3**/**8.0**	**21.5**/**7.5**	**25.7**/**7.4**	**22.4**/**7.2**	**19.0**/**6.2**	22.1/**8.0**	**24.1**/**8.6**	**21.6**/**7.7**

[a]Ave denotes the average values of ten datasets.

are less than six correct matched point pairs, we skip the current query image and set the NCC, NMI and SSIM values to 0.

In this part, we analyze the obtained PR and MS values of different registration methods. A significant fact is clarified here that as we have no knowledge about the true transformations, the PR and MS values are estimated and analyzed from a performance standpoint. Table 5 gives the mean values and standard deviations of estimated PR and MS values (only successfully aligned frames are considered). As presented in Table 5, in terms of PR values, the AKAZE feature method averagely performs 18.4%, 35.5% better than SIFT and SURF respectively, and is competitive with ORB method. In terms of MS values, the AKAZE scheme has 13.5%, 14.8%, and 1.9% higher values than SIFT, SURF, and ORB, respectively. Considering the performances of RANSAC method on different kinds of features in part A, it is evident that in the case of AOSLO image registration, the AKAZE feature is more robust, distinctive and repeatable compared to other three features.

Fig. 4 presents typical matching results obtained with four methods, and it is apparent that the AKAZE algorithm has the largest number of correct matches compared with other three methods. Moreover, the distribution of the AKAZE matches is relatively uniform, meaning that the information from different image areas is used to calculate the transformation matrix. Benefitted from the large number of correct matches and the relatively uniform distribution of these matches, the calculated transformation matrix is more accurate as well as robust. Conversely, it is obvious that although the number



(a) Total: 251 correct matches; Red-box: 66 correct matches

(b) Total: 161 correct matches; Red-box: 40 correct matches

(c) Total: 306 correct matches; Red-box: 18 correct matches

(d) Total: 380 correct matches; Red-box: 99 correct matches

FIGURE 4. Shown are feature matches found by different registration algorithms with RANSAC method. The individual lines connect matched features in the two images. (a) Inlier matches found by SIFT, (b) inlier matches found by SURF, (c) inlier matches found by ORB, (d) inlier matches found by AKAZE.

TABLE 6. The top half of the table presents the number of discontinuities that cannot be found by different registration methods. The bottom half of the table shows the number of continuities which are regarded as discontinuities by different methods. The smallest value in each column is highlighted in bold.

Correct Discontinuity											
	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	2	0	2	2	**0**	2	0	1	2	2	1.3
SURF	1	0	2	2	2	2	1	2	2	2	1.6
ORB	2	4	3	3	3	3	4	4	3	3	3.2
AKAZE	**0**	**0**	**1**	**1**	1	**1**	**0**	**0**	**2**	**1**	**0.7**
Correct Continuity											
	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	**1**	2	3	3	3	0	**0**	2	1	2	1.7
SURF	3	3	3	5	4	1	2	4	2	2	2.9
ORB	1	2	3	2	2	2	3	2	4	3	2.4
AKAZE	2	**1**	**2**	**1**	**2**	**0**	1	**0**	**1**	**2**	**1.2**

[a]Ave denotes the average values of ten datasets.

of inlier ORB matches is large, these matches are mostly concentrated on center areas of AOSLO images. Therefore the ORB scheme cannot obtain an accurate transformation matrix.

To better observe the evolution of MS and PR values in a dataset, we give the MS and PR values of one dataset in Fig. 5 (other datasets produce similar results, and not shown in this paper). In Fig. 5, the corresponding MS and PR values of all these registration methods rapidly decrease once there is a discontinue frame or a frame with large intra-frame distortions. In practice, these algorithms cannot identify all discontinue frames or consecutive frames. In Table 6, we summarize the number of discontinue frames and consecutive frames which cannot be found by these methods, and obviously, the AKAZE algorithm has the minimum number of error registrations.

We also provide the statistical results of NCC, NMI and SSIM values in Table 7, and it is easily seen that the AKAZE feature performs 2.3%, 3.2%, and 3.5% better than SIFT, SURF, and ORB features, respectively, in terms of NCC results. In terms of NMI values, the AKAZE feature has 0.4%, 0.5%, and 0.6% higher values than SIFT, SURF, and ORB features, respectively. In aspect of SSIM values, the AKAZE method obtains 2.2%, 3.0% and 3.1% higher values than SIFT, SURF and ORB respectively. In addition, as shown in Table 7, the AKAZE method not only obtains the highest mean values but also has very steady performance.

To better observe the evolution of NCC, NMI and SSIM values in a dataset, we provide the NCC, NMI and SSIM results of one dataset in Fig. 6 (other datasets produce similar results, and not shown in this paper). In Fig. 6, it is easily seen that for frames which are close to the reference

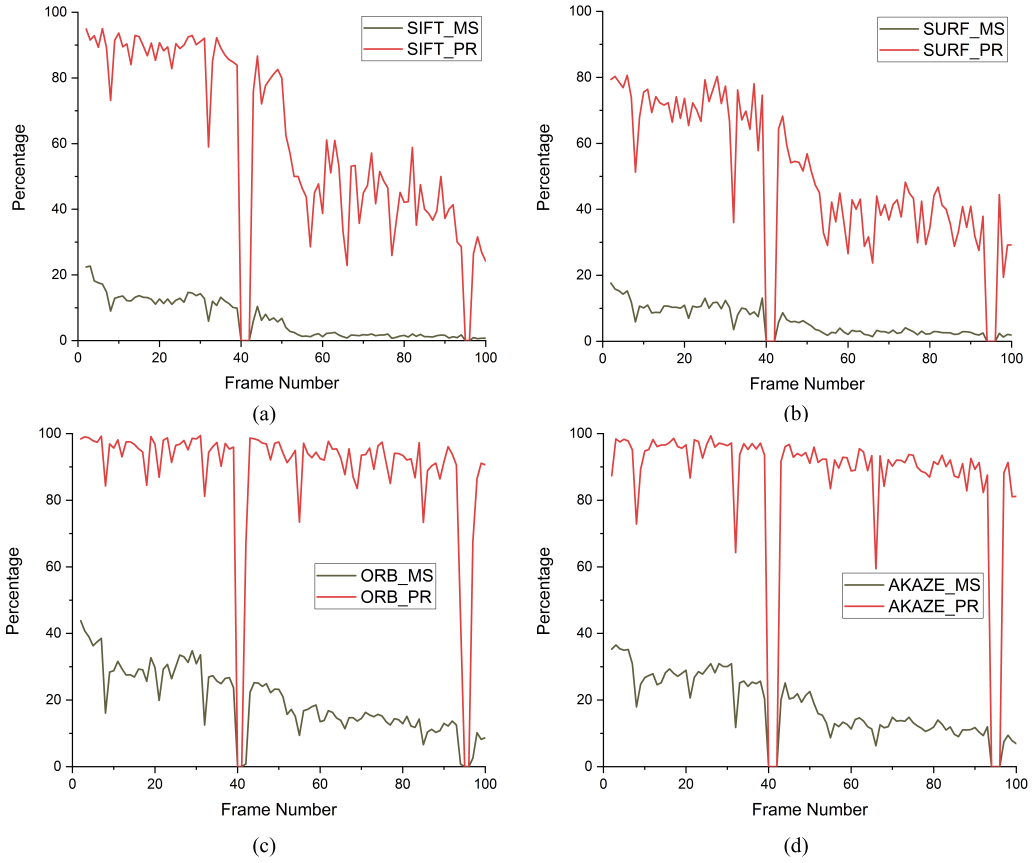


FIGURE 5. The evolution of PR and MS values of different registration methods in a dataset. (a) SIFT; (b) SURF; (c) ORB; (d) AKAZE.

TABLE 7. The mean values (Mean) and standard deviations (Std) of normalized cross-correlation (NCC), normalized mutual information (NMI) and structure similarity index (SSIM) values on ten datasets. The largest Mean and the smallest Std in each column are highlighted in bold.

NCC Value											
Mean/Std %	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	72/10.7	61/12.4	66/10.8	68/7.9	67/8.5	80/6.1	63/8.7	69/7.6	74/8.0	67/11.0	68.8/9.2
SURF	72/9.0	61/11.7	66/10.4	64/11.7	65/9.2	79/5.5	62/8.3	68/7.5	75/6.2	67/10.2	67.9/9.0
ORB	73/6.7	61/8.4	64/**9.1**	67/**6.5**	63/**7.7**	79/**5.1**	60/**7.0**	67/**6.4**	74/5.3	67/**8.2**	67.6/**7.0**
AKAZE	**76/6.5**	**67/7.1**	**68**/9.3	**69**/6.6	**68**/8.2	**80**/5.2	**64**/7.8	**70**/6.5	**77/5.0**	**70**/8.9	**71.1**/7.1
NMI Value											
Mean/Std %	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	10.7/2.0	6.1/1.8	6.3/1.5	7.6/1.4	6.7/1.4	12.7/1.3	6.0/1.3	7.9/1.4	11.5/1.9	7.7/1.8	8.3/1.6
SURF	10.6/1.8	6.1/1.7	6.2/1.5	7.1/1.8	6.4/1.5	12.6/1.2	5.9/1.3	7.7/1.4	11.6/1.5	7.6/1.7	8.2/1.5
ORB	10.7/**1.3**	6.0/1.3	5.8/**1.3**	7.5/**1.1**	6.0/**1.2**	12.4/**1.1**	5.7/**1.1**	7.7/**1.2**	11.0/**1.2**	7.5/**1.4**	8.1/**1.2**
AKAZE	**11.3/1.3**	**6.8/1.2**	**6.5/1.3**	**7.9**/1.2	**6.8**/1.3	**12.8**/1.2	**6.3**/1.2	**8.2/1.2**	**12.1**/1.3	**8.1**/1.5	**8.7**/1.3
SSIM Value											
Mean/Std %	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	37.5/9.9	37.0/10.1	41.4/8.8	40.3/8.0	43.5/8.1	43.7/6.7	36.0/8.0	39.6/7.2	38.9/9.0	42.5/9.7	40.0/8.5
SURF	36.8/9.4	37.0/9.6	41.0/8.7	37.1/10.9	41.6/8.7	43.2/6.3	35.5/7.7	38.6/7.1	39.4/7.7	42.1/9.2	39.2/8.5
ORB	37.9/9.9	36.8/10.1	39.6/8.8	39.8/8.0	40.3/8.1	42.5/6.7	34.7/8.0	38.9/7.2	38.3/9.0	42.0/9.7	39.1/8.5
AKAZE	**40.7/7.0**	**40.9/6.5**	**42.8/7.8**	**42.2/6.6**	**44.4/7.7**	**44.3/5.7**	**37.9/6.9**	**41.7/5.9**	**41.9/6/4**	**44.8/7.9**	**42.2/6.9**

[a]Ave denotes the average values of ten datasets.

frame, both SIFT and AKAZE methods achieve the best alignment and SURF method is competitive with SIFT and AKAZE, whereas ORB method performs worse than other three methods. However, with the query frame departing from the reference frame, the performances of SIFT and SURF methods rapidly decrease and is even worse than ORB algorithm does. Conversely, the AKAZE method is more robust and performs much better than other three methods. The reason is that the difference between the reference image and query image becomes large with the distance

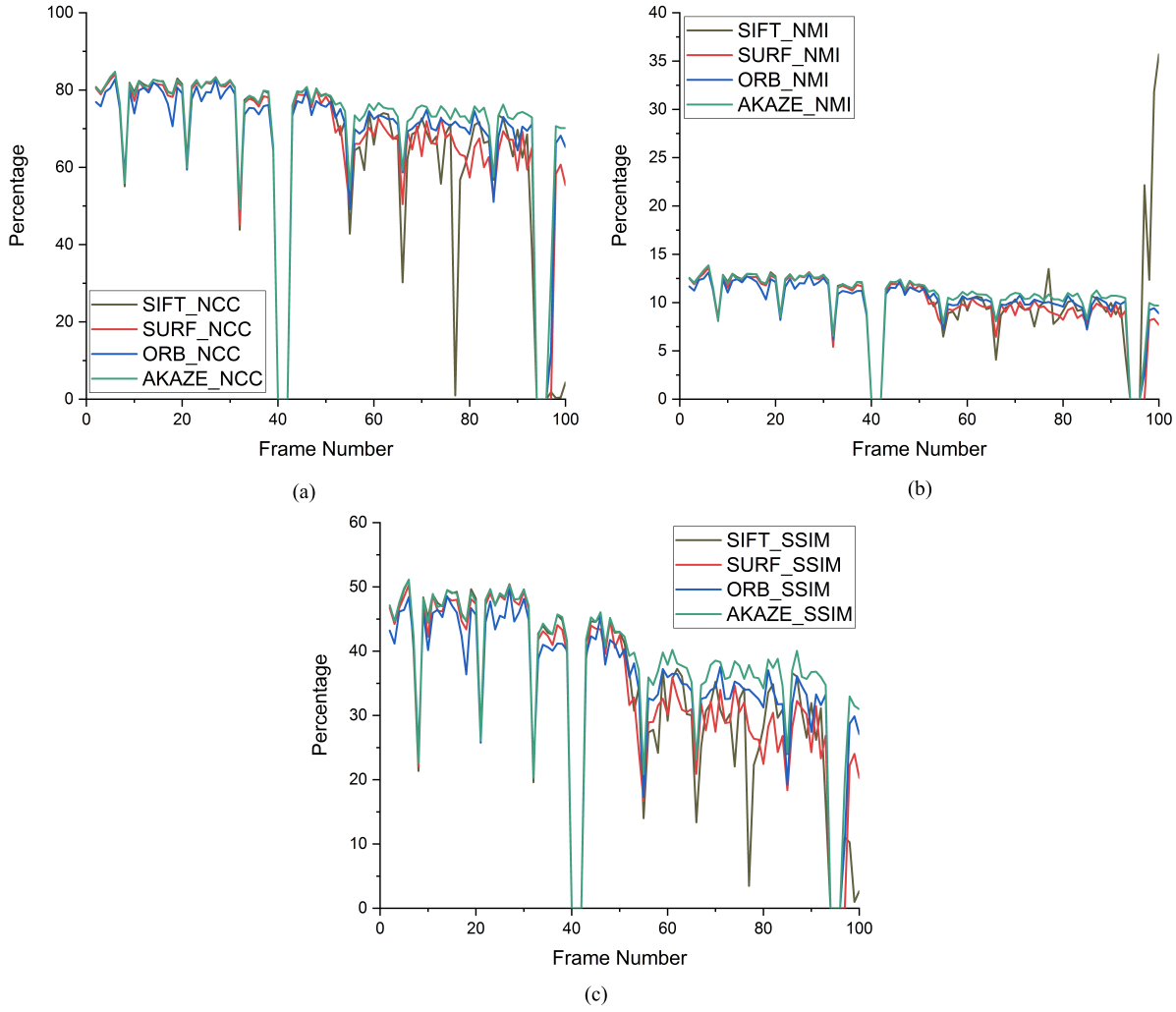


FIGURE 6. The evolution of NCC, NMI and SSIM values of different registration methods in a dataset. (a) NCC results, (b) NMI results and (c) SSIM results. The black, red, blue and green curves represent the results obtained by SIFT, SURF, ORB and AKAZE methods, respectively.

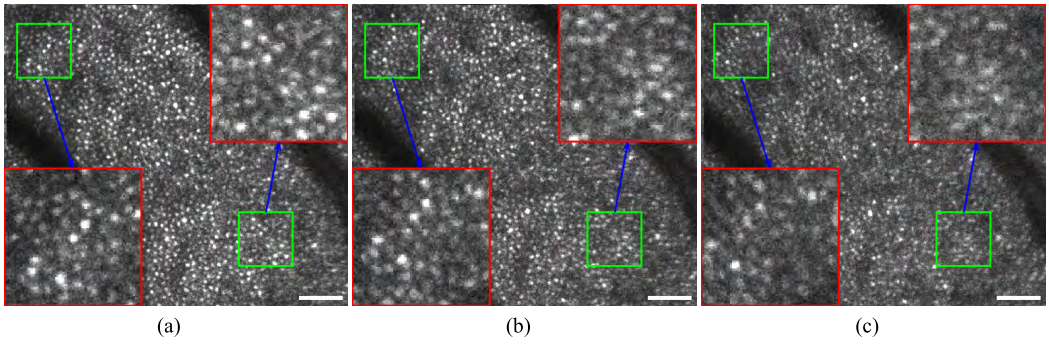


FIGURE 7. Frames exacted from different positions in a dataset. (a) The reference frame, (b) the 50-th frame and (c) the 100-th frame. The sub-images in green box were enlarged to red box with scale equaling 2.5. The scale bar is 50 μm.

between them increasing. Fig. 7 presents captured frames at different position of the dataset, and obvious differences between each other can be visible to anyone. Several factors (e.g. the variation of residual aberration, retina motion and light absorption) lead to the difference. Benefitted from the nonlinear diffusion filter, the AKAZE method is much more robust to the difference and feature points can be well located.

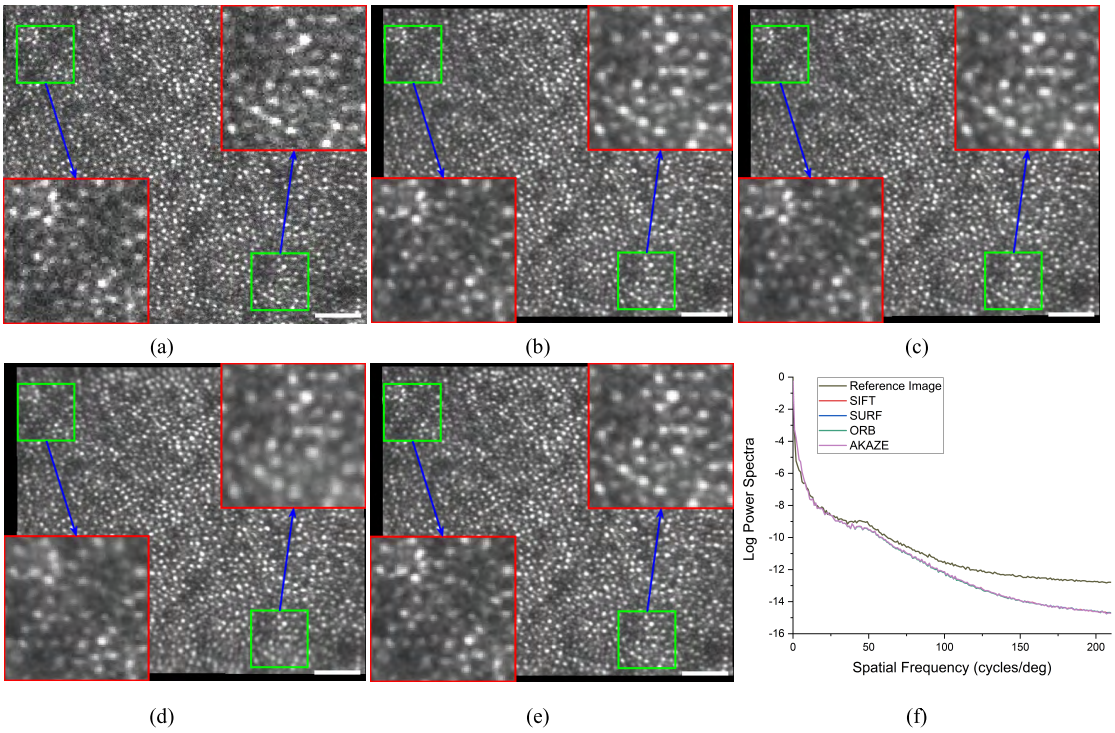


FIGURE 8. Images obtained by averaging 15 frames together with different registration methods. (a) Reference image acquired from a 24-year-old female with 1 ř FOV; (b) co-added image with the SIFT scheme (the image contrast is 0.4461); (c) co-added image with the SURF method (the image contrast is 0.4455); (d) co-added image with the ORB method (the image contrast is 0.4422); (e) co-added image with the AKAZE method (the image contrast is 0.4364); and (f) power spectra of these images, where the black, red, blue, green, and purple lines represent the power spectra of (a), (b), (c), (d), and (e), respectively. The sub-images in green box were enlarged to red box with scale equaling 2.5. The scale bar is 50 μm.

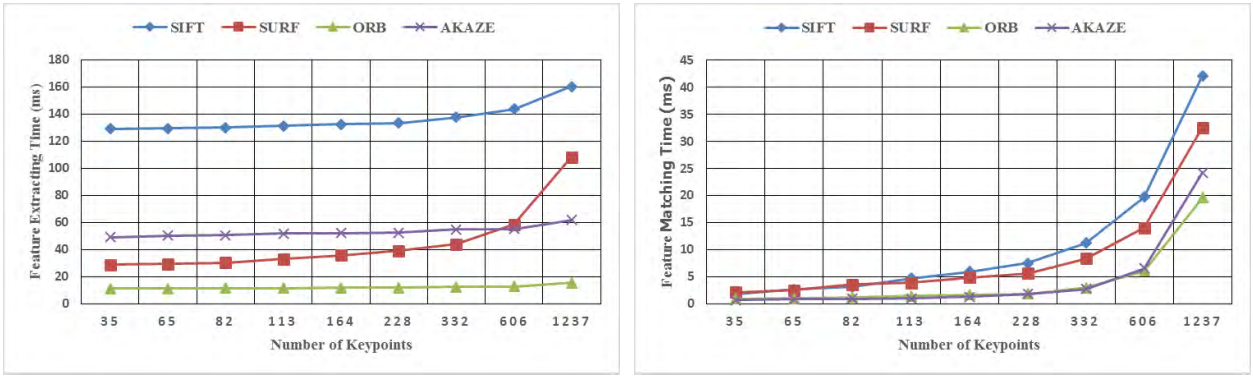


FIGURE 9. (a) Shows feature extracting time of different registration algorithms with different numbers of points detected in every image; (b) is the corresponding feature matching time.

Finally, 15 frames extracted from one dataset are matched by different registration methods and averaged, the results are shown in Fig. 8 (other datasets give similar results). If more frames are selected, the overlapping area will be small and the image quality is not improved obviously. The corresponding power spectra of these images is also presented in Fig. 8 (f). Table 8 presents the image contrasts of averaged images of ten datasets by co-adding 15 consecutive frames together, and it is obvious that the AKAZE method gets best contrast scores compared to other three registration methods.

C. RUNNING-TIME EVALUATION

In this section, we explore the running time of different methods. We take into account both the feature exaction (feature points detection and descriptors construction) and feature matching steps of different registration methods. To keep all algorithms having the same number of key points, relevant control parameters are changed according to the requirements of different detection methods. As shown in Fig. 9 (a), with the number of detected points increasing, generally, the feature extraction time increases for all methods, but it is

obviously that the SIFT and SURF algorithms are more sensitive to the number of detected points.

In Fig. 9, it is evidently visible that SIFT method is the most time-consuming method. In addition, although the feature extraction speed of the AKAZE scheme is slower than the SURF method does in cases where a small number of points are detected, it is still very competitive. In addition, the AKAZE algorithm becomes more efficient than SURF when more detected points are needed. Furthermore, although the ORB feature is the fastest method, as presented in above parts, its performance is not satisfying.

Moreover, as shown in Fig. 9 (b), in terms of feature matching, the AKAZE and ORB methods are faster than SIFT and SURF. When the number of feature points increases, the gaps are more visible. It is because of the difference in descriptor construction.

D. FRAME SELECTION AND THE MINIMUM NUMBER OF FRAMES FOR IMPROVING IMAGE QUALITY

Above results indicate that the AKAZE algorithm has better registration performance and faster speed compared to traditional GSS-based methods. Therefore, in this section, we use AKAZE algorithm to automatically accomplish AOSLO image dewarping. However, there are two problems needed to be solved. One problem is that we need to pick the ''good'' frames before superposing them. It means that the selected frames should have large overlapping regions, and the corresponding NCC, NMI and SSIM values should be as large as possible. To guarantee the real-time performance of the superposing procedure, the NCC, NMI, and SSIM cannot be used as criteria to select ''good'' frames. Therefore, selecting ''good'' frames is still a problem to be solved.

To tackle the problem, we first analyze the relationship between NCC values, PR values, and MS values. As presented in Fig. 10 (a), the NCC value is approximately proportional to the PR and MS values. It means that if a frame has a high NCC value, then there is a high probability that it also has high PR and MS values. Figure 10 (b) gives the relationship between NMI values, PR values, and MS values. Similarly, the NMI value is positively related to the PR and MS values. Therefore, we propose to select appropriate PR and MS values as thresholds to choose ''good'' frames. To find suitable thresholds, we firstly fit these data. In Fig. 10, it is clearly visible that to ensure the NCC and NMI values are larger than their mean values, the PR value should be larger than 0.833 and the MS value should not be smaller than 0.177. Therefore, in our experiment, we consider a query frame which has a PR value larger than 0.85 and MS value larger than 0.18 as a ''good'' frame.

The other problem to be solved is how many frames are needed to be superposed together to efficiently remove the distortions and noise. In practice, superposing a small number of frames cannot efficiently remove distortions and noise. On the other hand, with an increasing number of frames, the overlapping regions will become small, and when the number exceeds some value, the image quality will not be improved evidently. To decrease the running time and avoid unnecessary computation, we decide to explore the minimum number of frames needed to accomplish AOSLO image dewarping.

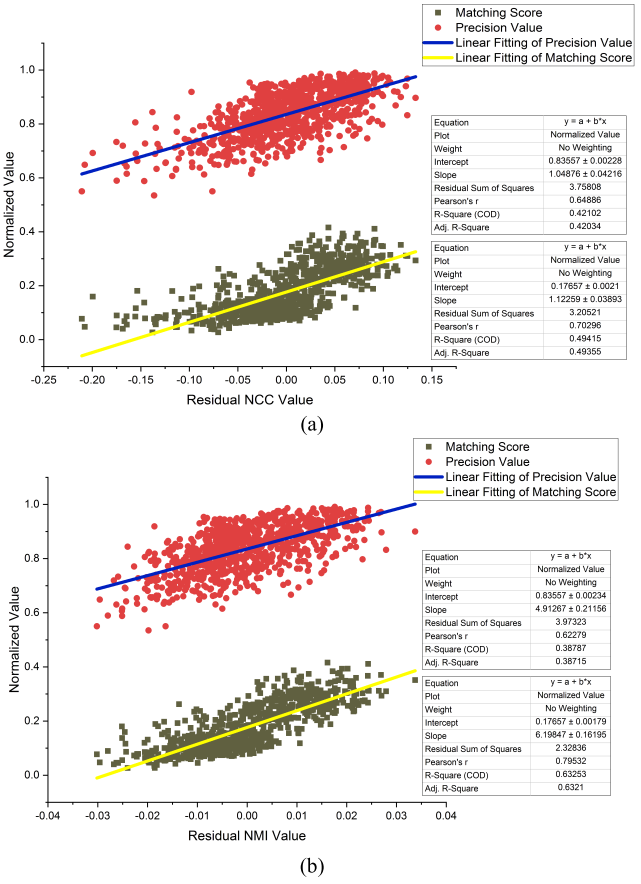


FIGURE 10. The relationship between NCC, NMI, PR, and MS values. (a) Empirical distribution of MS and PR values at different residual NCC values, where the blue and yellow lines represent the linear fitting of distributions of PR and MS values, respectively, the x axis denotes the regularized NCC values by subtracting the mean NCC values in each dataset; (b) empirical distribution of MS and PR values at different residual NMI values, where the blue and yellow lines represent the linear fitting of distributions of PR and MS values, respectively, the x axis denotes the regularized NMI values by subtracting the mean NMI values in each dataset.

We firstly select ''good'' frames from each dataset, using the method introduced above and obtain a series of averaged images by superposing different numbers of frames. Next, the power spectra of these averaged images is calculated. Finally, the minimum number of frames is determined if we have

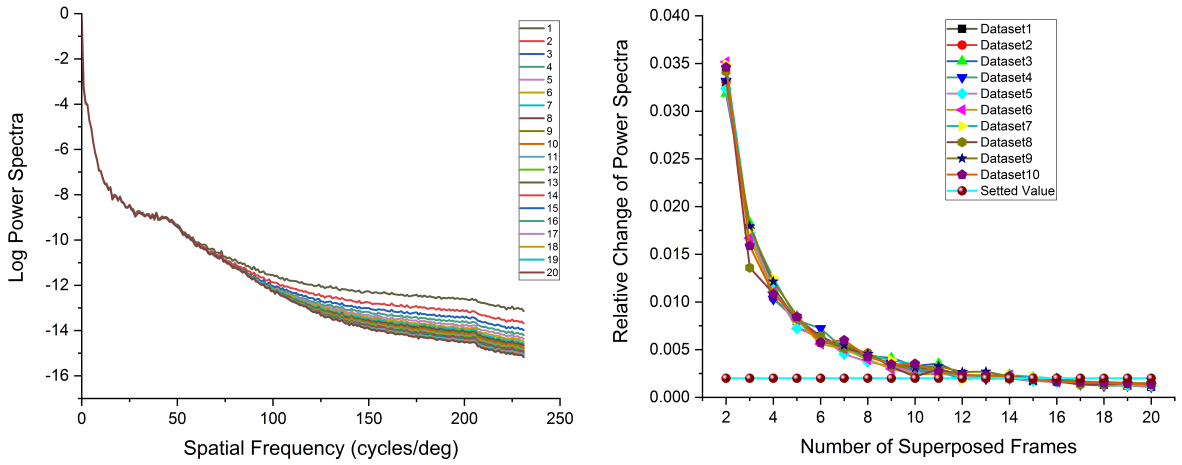
$$\Delta S^{i+1} = \frac{\left\| PS^{i+1} - PS^{i} \right\|_2}{\left\| PS^{i} \right\|_2} \leq c \quad 1 \leq i \leq N-1. \quad (13)$$

Here, PS^i denotes the logarithmic power spectrum of the superposed image that averages i frames together, c is a small positive constant (in our experiment we set it to 0.002), N indicates the total number of selected ''good'' frames, and

TABLE 8. The image contrasts of superposed images of ten datasets. Each image is obtained by co-adding 15 consecutive frames together. The largest contrast in each column is highlighted in bold.

	Image Contrast										
	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Ave[a]
SIFT	0.3890	0.2900	0.2932	0.3245	0.3123	0.4461	0.3402	0.3552	0.4309	0.3217	0.3503
SURF	0.3883	0.2891	0.2927	0.3245	0.3103	0.4455	0.3394	0.3546	0.4297	0.3204	0.3495
ORB	0.3836	0.2813	0.2859	0.3212	0.3024	0.4422	0.3368	0.3509	0.4282	0.3174	0.3450
AKAZE	**0.3897**	**0.2901**	**0.2939**	**0.3251**	**0.3128**	**0.4464**	**0.3416**	**0.3574**	**0.4318**	**0.3225**	**0.3511**

[a]Ave denotes the average values of ten datasets.



(a) (b)

FIGURE 11. The power spectrum of images. (a) gives the power spectrum of images which averaged different numbers of frames together in a dataset, (b) shows the relative changes of power spectrum of different datasets.

ΔS is the relative change in the power spectrum between two images.

We present the power spectrum of averaged images obtained by superposing different numbers of frames together from one dataset (other datasets produce similar results) in Fig. 11 (a), and Fig. 11 (b) presents the curves of different datasets. As seen in Fig. 11, when the number of frames exceeds ten, the image quality is not improved obviously. To suppress the value of c to 0.002, at least 15 images should be superposed together.

V. DISCCUSION

A. ALGORITHM EFFICIENCY

Shown in Table 2 and 5, the AKAZE algorithm achieves the best PR and MS values in most cases, indicating that the AKAZE feature has better distinctiveness and repeatability. In addition, the capability of discontinuity detection of four registration methods are analyzed in Table 6. Obviously, the AKAZE algorithm has the minimum number of error registrations. Besides the PR and MS metrics have been used to describe registration performance, the NCC, NMI and SSIM parameters of an overlapping area are also been adopted to character images alignment accuracy. As presented in Table 7, the AKAZE method not only obtains the highest NCC, NMI and SSIM values implying the best alignment accuracy, but also has the steadiest performance in almost all cases.

Presented in Table 8, the averaged images with the AKAZE method have the highest image contrasts indicating that the AKAZE method has the best dewarping performance compared to other three methods.

In Fig. 9, in aspect of feature extraction, it is clear that the speed of the AKAZE scheme is much faster than SIFT, and is very competitive with SURF. Especially, the AKAZE algorithm is faster than SURF when more detected points are needed to guarantee a high accuracy of registration. The ORB method is the fastest one of them, but it sacrifices the accuracy performance. In terms of feature matching, both AKAZE and ORB methods are faster than SIFT and SURF methods.

Therefore, the extensive experimental results above indicate that the AKAZE algorithm based on NSS is more suitable for AOSLO image dewarping, benefitted with a better accuracy and robustness, compared to GSS-based registration methods. In addition, the AKAZE method also has certain advantage in real-time performance.

B. ALGORITHM LIMITATION

Although the AKAZE scheme achieves better performance than several feature detection methods, it is far from perfect. There are at least two problems to be solved for AKAZE method. The first problem is that the AKAZE algorithm uses nonlinear diffusion filtering to build the NSS; although the FED framework sped up the procedure, its algorithmic

complexity limited its real-time performance. The other problem is that AKAZE method can detect more feature points than other registration methods, but matched points are relatively concentrated in the central region, which further affects the accuracy of the calculated transformation matrix.

C. ADVANTAGES OVER TRADITIONAL APPROACHES

In this work, we fully compared the performance of several registration methods on AOSLO image dewarping. It is evident that the AKAZE method based on NSS has better accuracy, robustness and real-time performance compared with traditional algorithms based on GSS. Then, we used the AKAZE algorithm to automatically remove AOSLO image warp. In addition, another notable point of this work is that we proposed a useful strategy to select ''good'' frames. Although the proposed method is based on the AKAZE feature, it can be applied to other registration methods with a similar operation. Last but not least, we proposed a flexible method to analyze the minimum number of frames needed to improve image quality. The experimental results demonstrated that at least 15 ''good'' frames is needed to improve the co-added image quality to some degree.

VI. CONCLUSIONS

In this paper, the AKAZE feature based on NSS was utilized to register AOSLO retinal images for the first time, and an efficient strategy based on matched feature points for frame selection was proposed to automatically accomplish AOSLO retinal image dewarping. Moreover, a flexible method based on power spectra analysis is proposed to study the minimum number of frames needed to improve image quality. Extensive experiments demonstrated that the AKAZE method is more suitable for AOSLO image dewarping benefitted with a better accuracy, stability and rapidity compared to several traditional registration methods based on GSS.

Several future directions can be taken to extend this work. Firstly, using multi-modal images simultaneously, including split-detection and dark-field images, could improve the quality of the averaged image [21], [24], [25]. in addition, improving the performance of the AKAZE method could provide faster speed and better alignment between images. For example, limiting the distribution of AKAZE features and making them more uniformly distributed is helpful to obtain a more accurate transformation matrix. Moreover, we can use a better gpu to run the AKAZE algorithm and improve its real-time performance [60]. finally, the method could be extended to address other tasks of retina image processing, such as wide-filed montaging and cone detection.

REFERENCES

[1] J. Liang, D. R. Williams, and D. T. Miller, ''Supernormal vision and high-resolution retinal imaging through adaptive optics,'' *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 14, pp. 2884–2892, Nov. 1997.
[2] B. Hermann *et al.*, ''Adaptive-optics ultrahigh-resolution optical coherence tomography,'' *Opt. Lett.*, vol. 29, pp. 2142–2144, Sep. 2004.
[3] A. Roorda, F. Romero-Borja, W. J. Donnelly, H. Queener, T. J. W. Hebert, and M. C. Campbell, ''Adaptive optics scanning laser ophthalmoscopy,'' *Opt. Express*, vol. 10, pp. 405–412, May 2002.
[4] R. J. Zawadzki *et al.*, ''Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal *in vivo* imaging,'' *Opt. Express*, vol. 13, pp. 8532–8546, Oct. 2005.
[5] Y. Zhang, J. Rha, R. S. Jonnal, and D. T. Miller, ''Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina,'' *Opt. Express*, vol. 13, pp. 4792–4811, Jun. 2005.
[6] D. X. Hammer, R. D. Ferguson, C. E. Bigelow, N. V. Iftimia, T. E. Ustun, and S. A. Burns, ''Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging,'' *Opt. Express*, vol. 14, pp. 3354–3367, Apr. 2006.
[7] J. Lu, H. Li, L. Wei, G. Shi, and Y. Zhang, ''Retina imaging *in vivo* with the adaptive optics confocal scanning laser ophthalmoscope,'' in *Proc. 8th Int. Conf. Photon. Imag. Biol. Med.*, Oct. 2009, Art. no. 75191I.
[8] P. Godara, A. M. Dubis, A. Roorda, J. L. Duncan, and J. Carroll, ''Adaptive optics retinal imaging: Emerging clinical applications,'' *Optometry Vis. Sci., Off. Publication Amer. Acad. Optometry*, vol. 87, no. 12, p. 930, Dec. 2010.
[9] M. Mujat, R. D. Ferguson, A. H. Patel, N. Iftimia, N. Lue, and D. X. Hammer, ''High resolution multimodal clinical ophthalmic imaging system,'' *Opt. Express*, vol. 18, pp. 11607–11621, May 2010.
[10] A. Roorda, ''Applications of adaptive optics scanning laser ophthalmoscopy,'' *Optometry Vis. Sci., Off. Publication Amer. Acad.*, vol. 87, p. 260, Apr. 2010.
[11] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, ''The role of fixational eye movements in visual perception,'' *Nature Rev. Neurosci.*, vol. 5, no. 3, p. 229, 2004.
[12] S. Martinez-Conde, ''Fixational eye movements in normal and pathological vision,'' *Prog. Brain Res.*, vol. 154, pp. 151–176, Sep. 2006.
[13] S. B. Stevenson and A. Roorda, ''Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy,'' *Proc. SPIE*, vol. 5688, pp. 145–151, May 2005.
[14] C. R. Vogel, D. W. Arathorn, A. Roorda, and A. Parker, ''Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy,'' *Opt. Express*, vol. 14, pp. 487–497, Jan. 2006.
[15] S. Faisan, D. Lara, and C. Paterson, ''Scanning ophthalmoscope retinal image registration using one-dimensional deformation fields,'' *Opt. Express*, vol. 19, pp. 4157–4169, Feb. 2011.
[16] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, ''A partial intensity invariant feature descriptor for multimodal retinal image registration,'' *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, Jul. 2010.
[17] H. Li, J. Lu, G. Shi, and Y. Zhang, ''Tracking features in retinal images of adaptive optics confocal scanning laser ophthalmoscope using KLT-SIFT algorithm,'' *Biomed. Opt. Express*, vol. 1, pp. 31–40, Aug. 2010.
[18] H. Li, H. Yang, G. Shi, and Y. Zhang, ''Adaptive optics retinal image registration from scale-invariant feature transform,'' *Optik*, vol. 122, pp. 839–841, May 2011.
[19] Z. Ghassabi, J. Shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, ''An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors,'' *EURASIP J. Image Video Process.*, vol. 1, no. 1, p. 25, Dec. 2013.
[20] H. Li, J. Lu, G. Shi, and Y. Zhang, ''Automatic montage of retinal images in adaptive optics confocal scanning laser ophthalmoscope,'' *Opt. Eng.*, vol. 51, no. 5, May 2012, Art. no. 057008.
[21] M. Chen, R. F. Cooper, G. K. Han, J. Gee, D. H. Brainard, and J. I. Morgan, ''Multi-modal automatic montaging of adaptive optics retinal images,'' *Biomed. Opt. Express*, vol. 7, p. 4899, Dec. 2016.
[22] Y. He *et al.*, ''Optimization of SIFT algorithm for fast-image feature extraction in line-scanning ophthalmoscope,'' *Optik*, vol. 152, pp. 21–28, Jan. 2018.
[23] H. Li, J. Lu, G. Shi, and Y. Zhang, ''Image registration in adaptive optics confocal scanning laser ophthalmoscope,'' *Amer. J. Eng. Technol. Res.*, vol. 15, pp. 42–46, Aug. 2015.
[24] B. Davidson *et al.*, ''Fast adaptive optics scanning light ophthalmoscope retinal montaging,'' *Biomed. Opt. Express*, vol. 9, pp. 4317–4328, Sep. 2018.
[25] W. Cao, F. Lyu, Z. He, G. Cao, and Z. He, ''Multimodal medical image registration based on feature spheres in geometric algebra,'' *IEEE Access*, vol. 6, pp. 21164–21172, 2018.
[26] S. Leutenegger, M. Chli, and R. Y. Siegwart, ''BRISK: Binary robust invariant scalable keypoints,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2011, pp. 2548–2555.
[27] P. Perona and J. Malik, ''Scale-space and edge detection using anisotropic diffusion,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[28] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1281–1298, Oct. 2011.
[29] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 214–227.
[30] S. Wang, H. You, and K. Fu, "BFSIFT: A novel method to find feature matches for SAR image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 649–653, Jul. 2012.
[31] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016.
[32] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38544–38555, 2018.
[33] F. Ye, Y. Su, H. Xiao, X. Zhao, and W. Min, "Remote sensing image registration using convolutional neural network features," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 232–236, Feb. 2018.
[34] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 421–432, Mar. 1998.
[35] S. Grewenig, J. Weickert, and A. Bruhn, *From Box Filtering to Fast Explicit Diffusion*. Berlin, Germany: Springer, 2010.
[36] J. Weickert, S. Grewenig, C. Schroers, and A. Bruhn, "Cyclic schemes for PDE-based image analysis," *Int. J. Comput. Vis.*, vol. 118, no. 3, pp. 275–299, 2016.
[37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[38] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, May 2006, pp. 404–417.
[39] X. Yang and K.-T. Cheng, "LDB: An ultra-fast feature for scalable augmented reality on mobile devices," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Nov. 2012, pp. 49–57.
[40] *American National Standard for Safe Use of Lasers*, Laser Inst. Amer., Washington, DC, USA,2007.
[41] T. Y. Chui, D. A. VanNasdale, and S. A. Burns, "The use of forward scatter to improve retinal vascular imaging with an adaptive optics scanning laser ophthalmoscope," *Biomed. Opt. Express*, vol. 3, pp. 2537–2549, Oct. 2012.
[42] A. Uji, S. Ooto, M. Hangai, S. Arichika, and N. Yoshimura, "Image quality improvement in adaptive optics scanning laser ophthalmoscopy assisted capillary visualization using B-spline-based elastic image registration," *PLoS ONE*, vol. 8, no. 11, 2013, Art. no. e80106.
[43] Y. He, Y. Wang, L. Wei, X. Li, J. Yang, and Y. Zhang, "Improving retinal image quality using registration with an SIFT algorithm in quasi-confocal line scanning ophthalmoscope," in *Proc. Oxygen Transp. Tissue*, Oct. 2017, pp. 183–190.
[44] J. Heinly, E. Dunn, and J. M. Frahm, "Comparative evaluation of binary features," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 759–773.
[45] S. Isik, "A comparative evaluation of well-known feature detectors and descriptors," *Int. J. Appl. Math., Electron. Comput.*, vol. 3, no. 1, pp. 1–6, 2014.
[46] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.
[47] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.
[48] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.
[49] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2385–2401, Nov. 2009.
[50] J. Wu, M. Kim, J. Peters, H. Chung, and S. S. Samant, "Evaluation of optimization methods for intensity-based 2D-3D registration in X-ray guided interventions," *Med. Phys.*, vol. 36, May 2009, Art. no. 796223.
[51] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.
[52] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[53] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[54] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 7, no. 10, pp. 2032–2040, 1990.
[55] J. Lu, B. Gu, X. Wang, and Y. Zhang, "High-speed adaptive optics line scan confocal retinal imaging for human eye," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0169358.
[56] B. Gu, X. Wang, M. D. Twa, J. Tam, C. A. Girkin, and Y. Zhang, "Noninvasive *in vivo* characterization of erythrocyte motion in human retinal capillaries using high-speed adaptive optics near-confocal imaging," *Biomed. Opt. Express*, vol. 9, pp. 3653–3677, Aug. 2018.
[57] L. Mariotti and N. Devaney, "Performance analysis of cone detection algorithms," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, pp. 497–506, Apr. 2015.
[58] X. Fei, J. Zhao, H. Zhao, D. Yun, and Y. Zhang, "Deblurring adaptive optics retinal images using deep convolutional neural networks," *Biomed. Opt. Express*, vol. 8, no. 12, pp. 5675–5687, 2017.
[59] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *ACM Commun.*, vol. 24, no. 6, pp. 381–395, 1981.
[60] K. Punithakumar, P. Boulanger, and M. Noga, "A GPU-accelerated deformable image registration algorithm with applications to right ventricular segmentation," *IEEE Access*, vol. 5, pp. 20374–20382, 2017.

Authors' photographs and biographies not available at the time of publication.