RetinaMatch: Efficient Template Matching of Retina Images for Teleophthalmology

Chen Gong , N. Benjamin Erichson, John P. Kelly, Laura Trutoiu, Brian T. Schowengerdt,
Steven L. Brunton, and Eric J. Seibel

Abstract— Retinal template matching and registration is an important challenge in teleophthalmology with low-cost imaging devices. However, the images from such devices generally have a small field of view (FOV) and image quality degradations, making matching difficult. In this paper, we develop an efficient and accurate retinal matching technique that combines dimension reduction and mutual information (MI), called RetinaMatch. The dimension reduction initializes the MI optimization as a coarse localization process, which narrows the optimization domain and avoids local optima. The effectiveness of RetinaMatch is demonstrated on the open fundus image database STARE with simulated reduced FOV and anticipated degradations, and on retinal images acquired by adapter-based optics attached to a smartphone. RetinaMatch achieves a success rate over 94% on human retinal images with the matched target registration errors below 2 pixels on average, excluding the observer variability, outperforming standard template matching solutions. In the application of measuring vessel diameter repeatedly, single pixel errors are expected. In addition, our method can be used in the process of image mosaicking with area-based registration, providing a robust approach when feature-based methods fail. To the best of our knowledge, this is the first template matching algorithm for retina images with small template images from unconstrained retinal areas. In the context of the emerging mixed reality market, we envision automated retinal image matching and registration methods as transformative for advanced teleophthalmology and long-term retinal monitoring.

Index Terms— Retina image template matching, teleophthalmology, dimension reduction, mutual information, health monitoring.

I. INTRODUCTION

TELEMEDICINE applications are emerging at a rapid pace due to innovations in hardware and software,

and changing attitudes of clinicians, providers and consumers. Teleophthalmology is an important component of telemedicine, and it is now arguably the standard of care in linking patients in remote areas to ophthalmologists. Recently, low-cost teleophthalmology has been facilitated by smartphone-based fundus imaging. In addition, the emerging virtual and mixed reality sector may enable new teleophthalmology scenarios for long-term eye imaging and monitoring. However, in the case of portable fundus photography, non-mydriatic image quality is more vulnerable to distortions, such as uneven illumination, noise, blur and low contrast [1]. In this paper, we address the challenging problem of automated retinal image matching and registration to enable future teleophthalmology applications.

A. Motivation

The eye provides a unique opportunity to image internal biological tissue in vivo and many diseases can be diagnosed and monitored through ocular imaging. For example, diabetic retinopathy is a common retinal complication associated with diabetes, causing microaneurysms, exudates and hemorrhages on the retina [2], [3]. Changes of retinal arteries and veins, as well as their ratios, can be indicators of hypertension [4]. The timely detection of these pathological changes via regular retinal screening and analysis is particularly important for early diagnosis and prevention.

High-quality fundus images of the retina are traditionally acquired in a laboratory setting with expensive and cumbersome equipments. Acquiring high-quality fundus images poses a significant challenge for patients in rural and other underserved areas who must overcome significant hurdles to receive regular checkups in the clinic. Visiting an ophthalmologist is often inconvenient for patients in the city as well. In contrast, emerging portable and low-cost fundus cameras allow fast, accessible imaging of the retina, albeit with a decrease in image quality. Using portable fundus cameras outside the clinic connects rural patients with their doctors [5], [6]. By daily retinal monitoring and trend analysis of the data, ocular disease may no longer be considered the silent disease, as early onset is likely to be detectable and even predicted [7].

A typical example of a portable fundus camera involves a clip-on lens adapter attached to a smart-phone system [5]. These consumer-grade optical devices have two main disadvantages: small FOV and lower image quality than lab-based fundus cameras. The FOVs of current clip-on lenses range
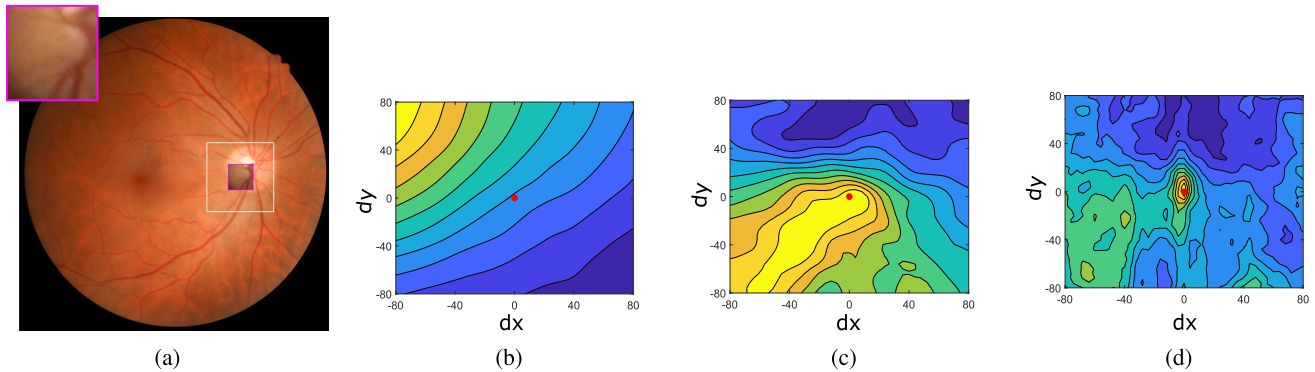
Fig. 1. Alignment functions with respect to translations between the template and the white boxed area. The full FOV image in (a) is taken with the fundus camera in the clinic. The top left image in the magenta square is the template captured by a typical adapter-based fundus camera *D-eye*, having only translations along two axes. (b)(c)(d) show the alignment function between the template and regions within the white boxed area. The true alignment position is (0, 0) – see red dots. Only NMI shows an obvious maximum at the alignment position. Note that the optimal value of SSD is minimum and NCC and NMI are maximums.

from 5° to 20° in undilated eyes [5], [8]. In this case, many small images captured in the undilated eye at different locations are necessary to obtain adequate retinal imaging. The same retinal locations need to be re-imaged and matched in order to monitor changes longitudinally over time. Accordingly, all of the captured small FOV images can be registered and compared to a stored wide FOV retinal image. This reference image is a baseline which can be stitched together by a series of small FOV images, or can leverage wider FOV images captured from a conventional ophthalmoscope. Taking the small FOV images as the templates to be matched, it is a template matching process, as shown in Fig. 1(a). The template only covers a small area on the retina, thus is unlikely to be affected much by the nonlinear deformation due to the non-planar eyeball surface. The location of the template in the full image may be represented by an affine linear transformation, i.e. including translation, rotation, shear, and scaling. This provides a mathematical framework to formulate template matching as an optimization problem.

As described above, an accurate template matching method to deal with small FOV and low quality template images is needed in teleophthalmology. Since the method will be implemented on portable devices, the efficiency of computation is also a driving requirement.

B. Related Work

Much of the foundational work on template matching of retinal images is based on more general image registration methods, which have been comprehensively studied in recent years. However, general retina registration methods focus on matching image pairs that both have a large FOV with local deformations or different image modalities. Many existing retinal template matching algorithms are limited to detecting specific objects from the image, where the template always contains a certain feature, such as the optic disc, exudate and artifacts [9]–[11].

Retinal image registration itself is challenging: the non-vascular surface of retina is homogeneous in healthy retinas, while exhibiting a variety of pathologies in unhealthy retinas [12]. Retina images captured by adapter-based optics provide less information and have low image quality, which further increases the difficulty of template matching. It is instructive to introduce current retina image registration methods which can be used for template matching and their feasibility in addressing our stated problem. Retina image registration approaches can be classified into area-based and feature-based methods. Feature-based methods optimize the correspondence between extracted salient objects in retina images [12]–[16]. Bifurcations, fovea, and the optic disc are common features used for retinal image registration. A small FOV template has little probability of containing specific landmarks on the retina, thus the fovea and optic disc are not applicable. Vascular bifurcations are more common, while similarly, the small number of bifurcations in the template cannot form the basis of a robust registration. Besides, the extraction of the vascular network in poor quality images is difficult. General feature point based approaches are also implemented in retina registration, such as SIFT-based [17], [18] and SURF-based methods [19], [20]. These approaches can register the images in complex scenarios and are computationally efficient. They assume the feature point pairs can be reliably detected and matched to estimate the transformation. Although feasible in most cases, the process can fail on low-quality retina images without enough distinct features.

Area-based approaches match the intensity differences of an image pair under a similarity measure, such as SSD (sum of squared differences) [21], CC (Cross-Correlation) [22] and MI (mutual information) [23], then optimize the similarity measure by searching in the transformation space. Avoiding pixel-level feature detection, such approaches are more robust to poor quality images than feature-based approaches. However, retina images with sparse features and similar backgrounds are likely to lead the optimization into local extrema. Fig. 1 shows an example of the area-based method with three similarity measures. The small template image is captured by the adapter-based *D-eye* optics which is registered onto a full fundus image. Both of the images are acquired by the same modality. SSD and normalized CC (NCC) do not have an obvious peak at the alignment position (0,0), giving

no clear information on the alignment quality. Normalized MI (NMI) shows a maximum at the alignment position, while it still has local extrema which can interfere with the global optimization. Besides, when the size difference between the template and full image is too large, registration with MI can be computationally prohibitive.

C. Contributions

In this paper, we present RetinaMatch, a new template matching method that overcomes the challenges posed by registering small FOV and low-quality retinal images onto a full image. This approach is an improvement over the area-based methods that only optimize the MI metric [23], since it achieves more accurate and robust template matching near the alignment position, as shown in Fig. 1.

The unique aspect of our approach is that we combine dimension reduction methods with the MI-based registration to reduce the sensitivity to local minima, while improving the matching efficiency. An overview of our novel template matching framework is shown in Fig. 2. Specifically, the principal component analysis (PCA) and block PCA are used to localize the template image coarsely, then the resulting displacement parameters are used to initialize the MI metric optimization. The initial parameters provided by the coarse localization are in the convergence domain of MI metric. In this way, the transformation search space for optimization is narrowed significantly. The PCA computation is accelerated with randomized methods [24]–[26], which improves the coarse localization efficiency. Both the use of PCA for coarse localization and the use of randomized methods for acceleration are unique methods of implementation. Further, we have carefully compared PCA against several other dimension reduction techniques, and we find that PCA offers the best tradeoff in simplicity, accuracy, and efficiency. Another contribution is that this paper proposes an efficient image mosaicking algorithm based on the image dimension reduction. It accelerates the matching of overlapped images among unordered data, especially in image mosaicking with area-based registration methods.

The proposed method is validated on the STARE retinal dataset [27] with synthetic deformations, and *in vivo* data captured by a low-cost (<US$400) adapter-based optical device *D-eye*. The performance of different dimension reduction techniques are also compared on the STARE dataset. RetinaMatch can find the correct mapping even when the image is of poor quality with non-distinct features, whereas other methods fail due to unstable feature detection and local extrema.

II. PRELIMINARIES

A. PCA for Location Estimation

Dimension reduction methods allow the construction of low-dimensional summaries, while eliminating redundancies and noise in the data. To estimate the template location in the 2d space, the full image dimension is redundant, thus we apply dimension reduction methods for the template coarse localization. In this section we describe the dimension reduction methods we use in this paper.

Generally, we can categorize dimension reduction techniques as either linear or nonlinear. The most prominent linear technique is principal component analysis (PCA), which dates back to the work of [28] and [29]. PCA is selected as the dimension reduction method in RetinaMatch since it is simple and versatile. Specifically, PCA forms a set of new variables as a weighted linear combination of the input variables. Consider a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d]$ of dimension $n \times d$, where n denotes the number of observations and d is the number of variables. Further, we assume that the matrix $\mathbf{X}$ is column-wise mean centered. The idea of PCA is to form a set of uncorrelated new variables (so called principal components) as a linear combination of the input variables:

$$\mathbf{z}_i = \mathbf{X}\mathbf{w}_i, \tag{1}$$

where $\mathbf{z}_i$ is the ith principal component (PC) and $\mathbf{w}_i$ is the weight vector. The first PC explains most of the variation in the data, the subsequent PCs then account for the remaining variation in descending order. Thereby, PCA imposes the constraint that the weight vectors are orthogonal. This problem can be expressed compactly as the following minimization:

$$\text{minimize } \|\mathbf{X} - \mathbf{Z}\mathbf{W}\|_F^2$$
$$\text{subject to } \mathbf{W}^\top\mathbf{W} = \mathbf{I}, \tag{2}$$

where $\|.\|_F$ is the Frobenius norm. The weight matrix $\mathbf{W}$ that maps the input data to a subspace turns out to be the right singular vectors of the input matrix $\mathbf{X}$. Often a low-rank approximation is desirable, i.e., we compute only the k dominant PCs using a truncated weight matrix $\mathbf{W}_k = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k]$.

PCA is generally computed by the singular value decomposition (SVD). Many algorithms have been developed to streamline the computation of the SVD and PCA for high-dimensional data that exhibits low-dimensional patterns [30]. In particular, tremendous strides have been made to accelerate the SVD and related computations using randomized methods for linear algebra [24]–[26]. Since we have demonstrated high performance with less than 20 principal components, the randomized SVD is used to compute the principal components, improving the efficiency in this retinal mapping application for mobile platforms. The randomized algorithm proceeds by forming a sketch $\mathbf{Y}$ of the input matrix

$$\mathbf{Y} = \mathbf{X}\mathbf{\Omega}, \tag{3}$$

where $\mathbf{\Omega}$ is a $d \times l$ random test matrix, say with independent and identically distributed random standard normal entries. Thus, the l columns of $\mathbf{Y}$ are formed as a randomly weighted linear combination of the columns of the input matrix, providing a basis for the column space of $\mathbf{X}$. Note, that l is chosen to be slightly larger than the desired number of principal components. Next, we form an orthonormal basis $\mathbf{Q}$ using the QR decomposition $\mathbf{Y} = \mathbf{Q}\mathbf{R}$. Now, we use this basis matrix to project the input data matrix to low-dimensional space

$$\mathbf{B} = \mathbf{Q}^\top\mathbf{X}. \tag{4}$$

This smaller matrix $\mathbf{B}$ of dimension $l \times d$ can then be used to efficiently compute the low-rank SVD and subsequently

the dominant principal components. Given the SVD of $\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}^\top$, we obtain the approximate principal components as

$$\mathbf{Z} = \mathbf{QU}\Sigma = \mathbf{XV}. \qquad (5)$$

Here, $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors and the diagonal elements of Σ are the corresponding singular values. The approximation accuracy can be controlled via additional oversampling and power iterations, for details see [31].

It is important to note that PCA is sensitive to outliers, occlusions, and corruption in the data. In ophthalmological imaging applications, there are several potential sources of corruption and outliers when imaging the full image, including blur, uncorrected astigmatism, inhomogeneous illumination, glare from crystalline lens opacity, internal reflections (e.g., from the vitreoretinal interface and lens), transient floaters in the vitreous, and shot noise in the camera. Further, there is always a trade-off between illumination and image quality, and there is strong motivation to introduce as little light as necessary for the patient comfort and health. The robust principal component analysis (RPCA) [32], [33] was introduced specifically to address this issue, decomposing a data matrix into the sum of a matrix containing low-rank coherent structure and a sparse matrix of outliers and corrupt entries. In general, RPCA is more expensive than PCA, requiring an iterative optimization to decompose the original matrix into sparse and low-rank components. Each step of the iteration is as expensive as regular PCA, and typically on the order of tens of iterations are required; however, PCA may be viewed as an offline step in our procedure, so that this additional computational cost is manageable. RPCA has been applied with success in retinal imaging applications to improve image quality [34], [35]. In the examples presented in this work, the data appears to have few enough outliers so that RPCA is not necessary, although it is important to keep RPCA as an option for data with outliers and corruption.

B. Mutual Information

In this section, we describe the maximization of MI for multimodal image registration. We define images $\mathbf{S}$ and $\widehat{\mathbf{S}}$ as the template and target images, respectively. A transform u is defined to map pixel locations $x \in \mathbf{S}$ to pixel locations in $\widehat{\mathbf{S}}$.

The main idea of the registration is to find a deformation $\widehat{u}$ at each pixel location x that maximizes the MI between the deformed template image $\mathbf{S}(u(x))$ and the target image $\widehat{\mathbf{S}}(\mathbf{x})$. Accordingly,

$$u_{opt} = \arg\min_{u} MI(\mathbf{S}(u(x)), \widehat{\mathbf{S}}(x)), \qquad (6)$$

where

$$MI(\mathbf{S}(u(x)), \widehat{\mathbf{S}}(x)) = \sum_{i_1 \in S}\sum_{i_2 \in \widehat{S}} p(i_1, i_2) log\left(\frac{p(i_1, i_2)}{p(i_1)p(i_2)}\right). \quad (7)$$

Here, i_1 and i_2 are the image intensity values in $\mathbf{S}(u(x))$ and $\widehat{\mathbf{S}}(x)$, respectively, and $p(i_1)$ and $p(i_2)$ are their marginal probability distributions while $p(i_1, i_2)$ is their joint probability distribution.

III. PROPOSED APPROACH

In this section, we describe RetinaMatch, which combines dimensionality reduction and mutual information based image registration. From Fig. 1 we can see MI performs better than other similarity metrics even on the same modality images, thus we focus on the MI criterion. Given a large FOV full image and a small FOV template image, our method can be used to localize the template on the full image accurately and efficiently. The full image can be a wide-field fundus image or a mosaicked one from D-eye images. The underlying concept is to use PCA and block PCA first for coarse localization, which can be a warm start to following accurate registration. In accurate registration, the MI metric is optimized to find the optimal transformation. Since the optimization domain has been narrowed to a small range near the optimal position with coarse localization, the accurate registration can achieve high accuracy and efficiency. Fig. 2 provides an overview of the general approach to RetinaMatch.

A. Coarse Localization With Dimension Reduction

We define the full image and the template as $\mathbf{F}$ and $\mathbf{S}$ respectively. The full image $\mathbf{F}$ is split into target images $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_N$:

$$\mathbf{I}_i = \phi(b_i, \mathbf{F}). \qquad (8)$$

The function ϕ crops $\mathbf{I}_i$ from $\mathbf{F}$ at b_i and $b_i = [x_i, y_i, h, w]$, where (x_i, y_i) denotes the center position and (h, w) denotes the width and height of the source image. There is a certain displacement f of neighboring target images in the x and y axes. As shown in Fig. 2(a), each target image has a large overlap with its neighbors. The overlap forms the redundancy of the data which can indicate the location distribution between each image and its neighbors. Applying dimension reduction techniques on such data we can obtain the low-dimensional distribution map of all target images.

Target images are resized to vectors and form the matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$. We obtain the low-dimensional distribution representation of the target image distribution by implementing PCA on $\mathbf{X}$:

$$\mathbf{Z} = \mathbf{XW}, \qquad (9)$$

where $\mathbf{Z} = [z_1, z_2, z_3, \ldots, z_N]^T \in \mathbb{R}^{n \times l}$, $\mathbf{W} \in \mathbb{R}^{d \times l}$ and $l \ll d$. The image space Ω_1 is mapped to a low-dimensional space Ω_2 with the mapping $\mathbf{W}$. $\mathbf{W}$ and $\mathbf{Z}$ are saved in the dictionary $\mathcal{D}$.

Given a template $\mathbf{S}$, the coarse position can be estimated by recognizing its nearest target image. The nearest target image in Ω_1 should also be the nearest representation of $\mathbf{S}$ in Ω_2. Accordingly, we obtain the low-dimension feature z_s of the template in Ω_2:

$$z_s = \tilde{s}\mathbf{W}, \qquad (10)$$

where $\tilde{s} \in \mathbb{R}^d$ is the reshaped vector of template $\mathbf{S}$. Let $\Delta(z_s, z)$ be the Euclidean distance between z_s and a feature z in $\mathbf{Z}$. z^* is the nearest target feature of the source image $\mathbf{S}$ in Ω_2:

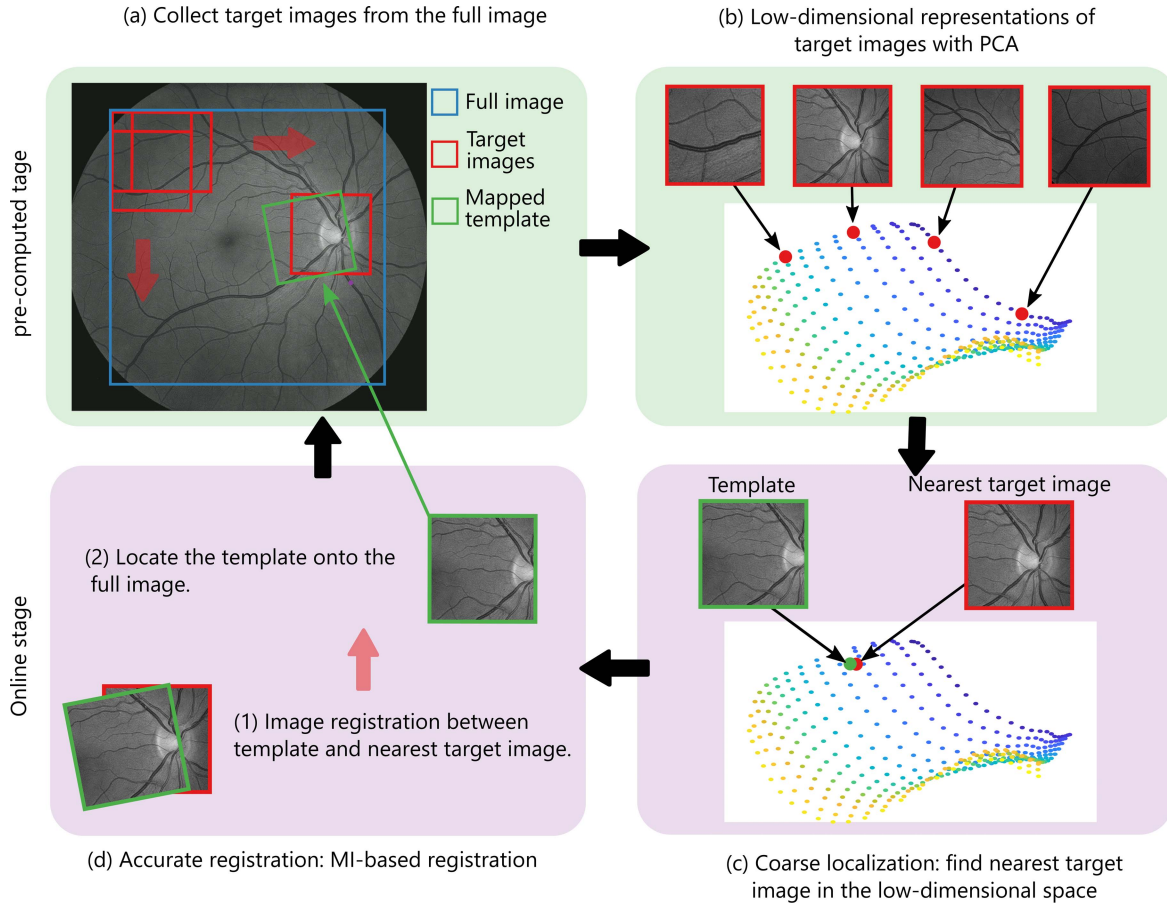$$z^* = \arg\min_{z} \Delta(z_s, z). \qquad (11)$$

Fig. 2. Schematic of the proposed retinal template matching method shown in four panels from (a) to (d). In panel (a) the wide-FOV full image is sampled with many overlapping target images. (b) Each target image is mapped into the low-dimensional space according to its positional relationship. (c) An example template is also mapped into this space and its nearest target image is found. (d) The nearest target image is registered with MI. The panels (a) and (b) in green can be pre-processed offline when the full image is obtained, while panels (c) and (d) are considered as the online stage. The schematic describes the method without using the improvement of block PCA. Please see Sect. III for more details of block PCA.

The corresponding target image location is used as the coarse location of $\mathbf{S}$. Ideally, the difference between the coarse location and the ground truth in x and y axes should be less than $f/2$ pixels.

In the first experiment in Sect. IV, PCA outperforms other non-linear dimension reduction methods, while the error is larger than $f/2$. The main reason is that the image degradation creates spurious features that contribute to the final classification. To reduce the influence of local spurious features, we implement block PCA to further improve the accuracy of the coarse localization. By computing the PCA of different local patches in the template, the effect of local features, which leading to the template can not be located correctly, is reduced.

Here we introduce the detailed process of the block PCA. Obtaining the nearest target image, we crop a larger image at the same position from the full image as the new target image $\mathbf{I}$. In this way, the template can have more overlap with the new target image when there is a large offset between two images. We segment $\mathbf{I}$ and the template $\mathbf{S}$ into small patches with the function $\tilde{\phi}$, where the patch size is smaller than the source image with the axial displacement of neighboring patches f'. Similarly, all image patches from $\mathbf{I}$ are mapped into the low-dimension space Ω_3 with $\mathbf{W}'$. Let $\mathbf{Z}'$ denote the low-dimensional representation of the target

image distribution. Each template patch is then mapped to the space with $\mathbf{W}'$.[1] The nearest target patch for each template patch is determined with the Euclidean distance as described before. The coordinates of each target patch represent the location of the mapped patch. We use the same weight for each region of the template for localization, thus the average of all template patches location can be taken as the template's location. Let b_m be the mean of the coordinates of selected nearest target patches, which then represents the center of the template on $\mathbf{I}$. Accordingly, the template location on the full image can be estimated and the region is cropped as the image $\widehat{\mathbf{S}}$. The accurate registration is then applied to the template $\mathbf{S}$ and image $\widehat{\mathbf{S}}$. In this way, the coarse localization provides an estimate of a good initial point for the accurate registration.

In the implementation of the proposed coarse localization, the full image is assumed to exist so the dictionary $\mathcal{D}$ can be built in advance. This is the pre-computed part as shown in Fig. 2 (a-b). The process after the template being acquired is called the online stage, involving the block PCA for coarse

---

[1]Fig. S1 in the supplementary material gives an example of the image patch mapping.

localization followed by the accurate registration. The online stage of the coarse localization is shown in Algorithm 1.

Algorithm 1: Coarse Localization: Online Stage

1 Map template $\mathbf{S}$ into space Ω_2: $z_s = \tilde{s}\mathbf{W}$.
2 Determine closest target image $\mathbf{I}$ with corresponding z^*: $z^* = \arg\min_z \Delta(z_s, z)$. $z^* \in \mathbf{Z}$.
3 Segment $\mathbf{S}$ into $[S_p^1, S_p^2, \ldots, S_p^n]$: $S_p^i = \tilde{\phi}(b_i, \mathbf{S})$; Segment $\mathbf{I}$ into $[I_p^1, I_p^2, \ldots, I_p^n]$: $I_p^i = \tilde{\phi}(b_i, \mathbf{I})$.
4 Map target patches I_p^i into space Ω_3: $\mathbf{Z}' = \mathbf{I_p}\mathbf{W}'$, where $\mathbf{I_p}$ is formed with vectorized I_p^i.
5 For each template patch S_p^i:
6   (i)Map S_p^i into space Ω_3: $\tilde{z_s}^i = S_p^i\mathbf{W}'$.
7   (ii)Determine its closest target patch $I_p^{Idx(i)}$ with index $Idx(i)$.
8 $b_m = \frac{1}{n}\sum_{i=1}^{n} b_{Idx(i)}$, where $b_{Idx(i)}$ is the coordinate of selected target patch $I_p^{Idx(i)}$.
9 **return** localization region $\widehat{S} = \phi(b_m, \mathbf{F})$).

B. Accurate Registration

In this section, images $\mathbf{S}$ and $\widehat{\mathbf{S}}$ are accurately registered with maximization of mutual information. The location of $\widehat{\mathbf{S}}$ on the full image $\mathbf{F}$ becomes the estimated displacement of the template $\mathbf{S}$. As the small FOV of template images, the relationship between the template and the full image can be modeled by linear transformations. In our work, the transform u for alignment is given as an affine transformation:

$$u = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 1 & 1 & 0 \end{bmatrix}. \tag{12}$$

From the MI equation 7, we can see the MI function has a discrete formulation which is not differentiable. Several solutions therefore are proposed to smooth the MI function to compute the MI derivatives and keep its accuracy. We use the method described in [36], where the joint probability distribution between the images $\mathbf{S}$ and $\widehat{\mathbf{S}}$ is estimated with a Parzen window.

The optimizer used for the MI maximization is based on Newton's method. The MI function is a quasi-concave function (Fig. 1(d)), and the parabolic hypothesis of the Newton's method is only valid near the convergence. When the initial transformation is on the convex part of the cost function, it will cause the optimization to diverge. In the example of Fig. 1(d), the normalized MI measure has local extrema interference. The proposed coarse localization provides a good initialization of the displacement for subsequent optimization of the MI alignment function. In the figure, the estimated alignment position is $(11, 9)$. The estimation is close to the optimal value and falls in the convex domain of the MI metric, which provides more efficient optimization and avoid local extrema.

After registration between images $\mathbf{S}$ and $\widehat{\mathbf{S}}$, the template $\mathbf{S}$ can be matched on the full image $\mathbf{F}$ based on the position

b_m of the selected region $\widehat{\mathbf{S}}$. Fig. 3 shows a schematic of the coarse localization process and intermediate results.

C. Image Stitching

As pointed out in Sect. I, the full retina image can be stitched into a panorama by using many small templates. Users must capture a series of images in naturally unconstrained eye positions to explore different regions of the retina. It is problematic to determine adjacent images before the registration when we apply area-based registration approaches, since they do not have effective descriptors for matching.

Related to the dimension reduction in the proposed template matching method, here we present Algorithm 2 to learn the positional relationship of images to be stitched. In this way, the adjacent images can be recognized and registered efficiently. For a series of small images $\mathbf{X_i}$, we form the matrix $\mathbf{X}$, as with the matrix $\mathbf{T}$. PCA is applied to $\mathbf{X}$ and returns the low-dimensional features for each image in Ω_2. The distance between features in Ω_2 indicates the distance between images. The nearest neighbor $\mathbf{X_j}$ of image $\mathbf{X_i}$ is the one with largest overlap; the image pair is then registered with MI-based approach. To improve the algorithm robustness, the 3-nearest neighbors for each image are first selected to compute MI with, and we keep the one with the largest metric value.

Algorithm 2: Image Stitching

1 Map images into space Ω_2: $\mathbf{Z} = \mathbf{X}\mathbf{W}$.
2 For each image $\mathbf{X_i}$:
3   (i).Find the nearest 3 neighbors $\mathbf{X_j}$ minimizing the feature distance $\Delta(\mathbf{Z_i}, \mathbf{Z_j})$.
4   (ii).Compute the Mutual Information between each $\mathbf{X_j}$ and $\mathbf{X_i}$ and take the adjacent image with highest MI.
5 Panorama $\mathbf{R}$ Mosaicking: Align all the adjacent images with mutual information based registration method.
6 Panorama blending.
7 **return** panorama $\mathbf{R}$.

IV. EXPERIMENTS

We present the performance of our template matching method on three experiments using retina images. For comparison, we use the global MI algorithm described in Mattes et al. [37] and ASIFT (modified SIFT for affine deformation) [38]. In the first experiment, each template is extracted from the full fundus image in the STARE dataset and matched back to it. The intermediate results of the coarse localization are also evaluated. In the second experiment, the template captured by the adapter-based optics is matched to the full fundus image captured by the clinical fundus camera. In the third experiment, a panorama is mosaicked from small templates first, and subsequently individual templates are matched to the panorama.

A. Fundus Images From STARE Dataset

In this experiment, we validate our method on simulated fundus images. We use images from the STARE dataset [27],
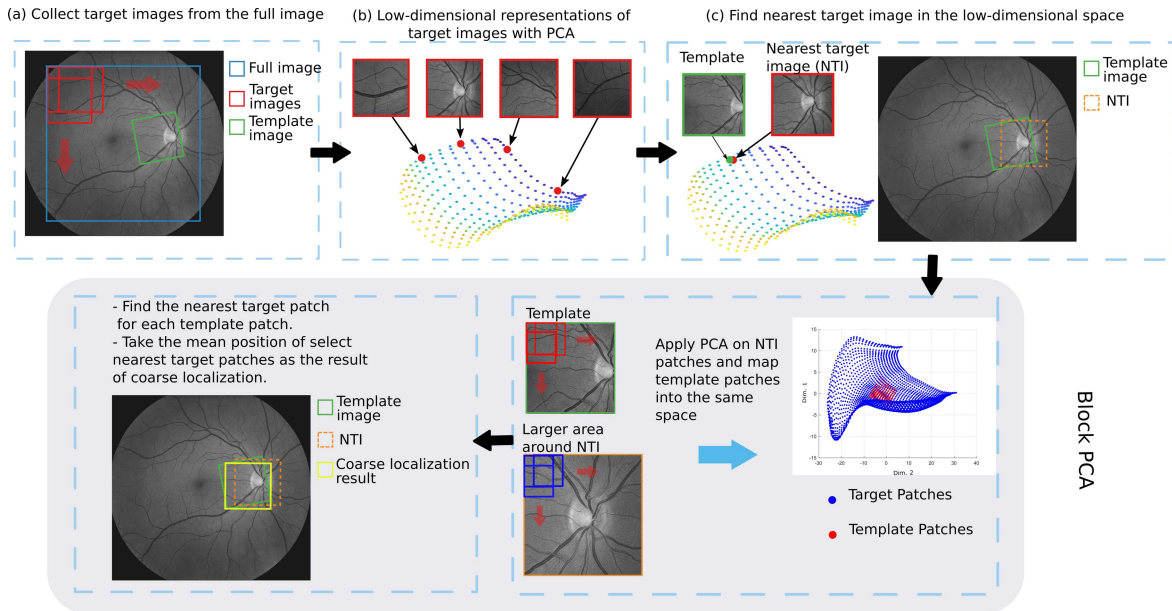
Fig. 3. Schematic of the coarse localization process: PCA and block PCA.

which consists of 400 raw fundus images of healthy and diseased retinas. Matching image pairs are simulated from this dataset. Each image pair includes a full fundus image selected randomly from the dataset and an affine transformation is applied to map it from a square into a parallelogram. The area within the mapped square is then cropped and warped (with the inverse affine transformation) to obtain the square template. The FOV of the template images is around 12° with a size of 200×200 pixels. The template dimension is 10% of the full image.

The ground truth is available in this experiment, thus root-mean-square (RMS) errors between corrected displacements and ground truth positions are used as a metric to measure the RetinaMatch accuracy. To evaluate the coarse localization, we take the center point distance between the template and the chosen target region.

1) Validation of the Coarse Localization: First the coarse localization with and without the block PCA refinement are tested. In the implementation, target images are generated with a displacement of $f = 10$ pixels and $f' = 5$ for the block PCA. We use the top 20 and 10 PCA features in the first PCA step and the block PCA respectively. The parameters are fixed in remaining experiments. Additionally, we test the coarse localization with two other non-linear dimension reduction methods: kernel PCA [39] and Isomap [40]. We compare the non-linear dimension reduction methods to see if the non-planar retina surface and the affine transformation affect the performance of the PCA-based linear method. In the kernel PCA, we compared Gaussian kernel and polynomial kernels with different degrees. The Gaussian kernel has better performance and is thus selected for kernel PCA. There may be other better kernels that better separate the data. However, finding this embedding space is labor-intensive and may need to be re-tuned for new image types, whereas PCA is more generic. The experiment is performed over 100 matching

TABLE I
COMPARISON OF COARSE LOCALIZATION WITH DIFFERENT
DIMENSION REDUCTION METHODS

	Mean errors	Success rate	Runtime
Kernel PCA	57	83%	0.7035s
Isomap	27	94%	2.3634s
PCA	14	100%	0.0065s
Block PCA	8	100%	0.6143s

image pairs created from the STARE dataset. The pixel-level errors (coarse localization error as described), success rates, and average runtimes of these methods are shown in Table I. The criterion of successful matches in the coarse localization is a pixel-level error of less than 40 pixels. It is verified that the PCA based coarse localization is more efficient, accurate and interpretable. Block PCA further improves the accuracy while the time spent is higher than PCA-only method. To further improve the online efficiency, the target patches mapping can be precomputed for each target images. The average time spent in this case will decrease to 0.0975s.

Additional experiments were carried out to test the proposed coarse localization under different image degradations. Five degradation types in five levels are considered as follows (images are in double format $\in [0, 1]$): affine transform with the rotation/shear parameter of {5°/0.1,10°/ 0.2,15°/0.2,15°/0.3,20°/0.3}; additive Gaussian noise with standard deviation varied from 10% to 50% of the pixel value; image blurring with Gaussian kernels with standard deviation of {0.5,1,1.5,2,2.5}; intensity changes of {4%,8%,12%,16%,20%} of graylevels in the image, which is the nonlinear brightness change; add artificial pathological features of 1-5 levels with increasing amount and size, such as the spot of exudate (bright spots), hemorrhage (dark spots)
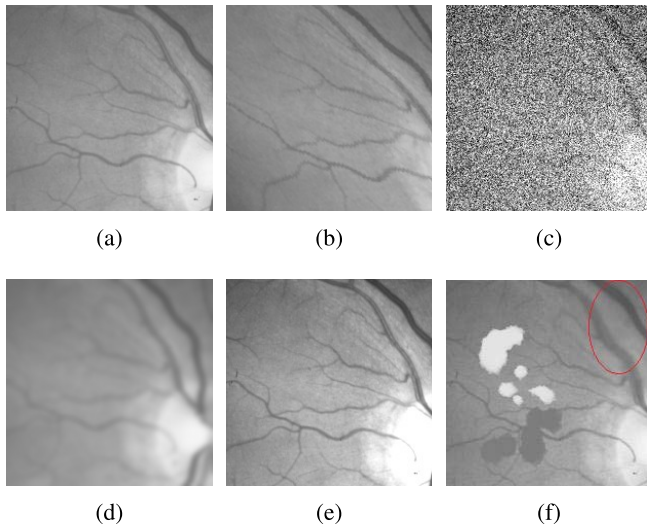
Fig. 4. Examples of highest level degradations in each sequence. Please note that (b) is the template generated with affine deformation thus the image content is not the same. In the artificial features (f), the bright and dark spots simulate the exudate and hemorrhage, respectively. The width of vessels in the circled area is enlarged. (a) Normal. (b) Affine deformation. (c) Gaussian noise. (d) Blur. (e) Brightness. (f) Artificial features.

TABLE II
SUCCESS RATES OF COARSE LOCALIZATION PER
DEGRADATION LEVEL

Distortion level	1	2	3	4	5
Affine deformation	100%	99%	95%	81%	75%
Noise	100%	100%	100%	100%	100%
Blur	100%	100%	100%	100%	100%
Brightness change	100%	100%	100%	100%	97%
Artificial features	100%	100%	100%	100%	98%

and vessel width changes (enlarge/shrink vessel regions). Fig. 4 provides examples of the highest level degradation in each sequence. For each sequence and degradation level, we create 100 matching image pairs as described above. All degradations are applied to the template in each pair. Fig. S2[2] shows template examples of the highest degradation in each sequence. The coarse localization achieves high success rates across the dataset in different degradations, with the exception of the highest level of linear deformation sequence. However, the limitation to smaller eye rotation angle is physiologically based. The human eye has a limited range of torsional rotation with respect to the visual axis [41]. Checking a large set of data, we find the real affine deformation in the adapter-based optics imaging is less than level three (rotation/shear: 15°/0.2).

2) Validation of the Template Matching: We examine RetinaMatch's final performance under the same sequences and degradations described earlier, but with two additional template matching approaches: feature-based ASIFT and global MI registration. The success rate of different methods are presented in Fig. 5, where the successful matches are that

[2]Fig. S2 is in the supplementary material.

the RMS error is less than 8 pixels. The accuracy of ASIFT decreases significantly at higher degradation levels of noise, blur, and artifacts, due to feature-point instability. The global MI registration method cannot always converge to the correct affine transformation parameters using such small templates, thus it has a low success rate even without degradations. The performance further declines in high-level degradations of artifacts and affine. RetinaMatch has a success rate of 100% in most sequences and degradation levels, except the high-level affine deformation. As described above, the real-world affine deformation would be less than level 3. The improvement of RetinaMatch efficiency over global MI depends on the size difference between two matched images. In this experiment, the average runtime of RetinaMatch is around 50% less than that of global MI, since it narrows the search domain of the MI optimization significantly. The computation of feature points in ASIFT is expensive and it takes four times longer than RetinaMatch. ASIFT is selected for comparison because it can generate more robust feature points than SIFT.

B. In Vivo D-eye Data and Full Fundus Image

D-eye is a typical adapter-based optical system which can convert the digital camera on smartphones to a fundus camera (https://www.d-eyecare.com/). Fig. 6 show several examples of D-eye images. The relatively small FOV of D-eye can be useful to monitor the retinal health over time with comparison to a wide FOV baseline image taken at the ophthalmology clinic. With our algorithmic approach, the captured D-eye images can be matched onto the full image for automatic comparison. The latest data with retina changes can also replace the original area on the full image, maintaining a record of longitudinal changes. In this way, it offers the opportunity for a quick overall retina analysis outside the clinic, with automatic diagnostic approaches such as described in [42] and [43].

This experiment is a case study with a series of D-eye data captured from one person with a healthy retina. We converted the iphone 6 to a fundus camera with D-eye, then collected the data in a dim room to provide a larger pupil and proper image contrast. The eyeball was free to rotate which allowed us to obtain images covering different regions of the retina. The collected *D-eye* images have an average FOV of 4° and a resolution of about 50 pixels/degree. The full fundus image is taken with a Kowa Nonmyd alpha-D III retinal camera, as shown in Fig. 1(a). It has a 45° FOV with a resolution of 75 pixels/degree. The D-eye images are around 0.7% of the full image. Captured with different devices, the brightness and contrast varies greatly between the image pair to be matched.

We first validate our method by matching 100 *D-eye* template images onto the full image. The ASIFT and global MI methods are also implemented. Additionally, we add pathological artificial features on the 100 D-eye templates to test the algorithm robustness to retina pathological changes. The accuracy of the template matching is evaluated using target registration error (TRE) [44]. For each template, four corresponding landmarks are selected by an trained observer, two trained observers then selected the corresponding landmarks
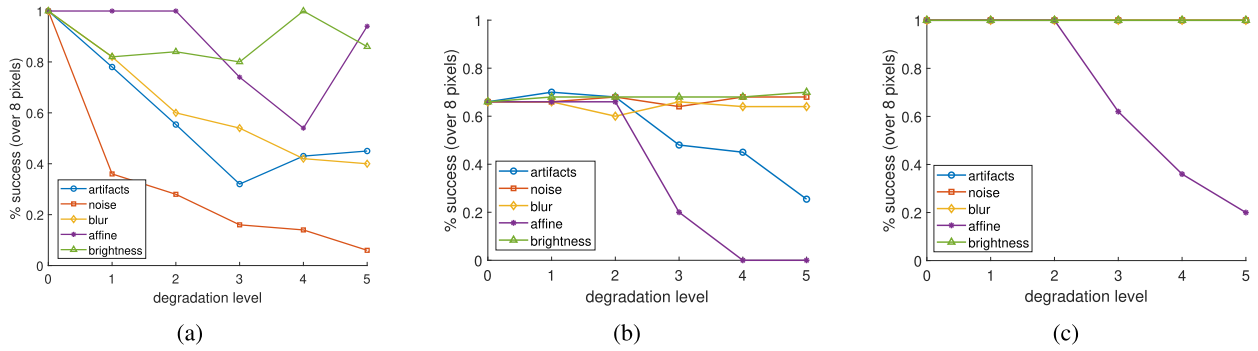
(a) (b) (c)

Fig. 5. Performance of template matching methods under different image degradations. In each, the x-axis stands for the increasing levels of image degradation, ranging from 0 (no degradation) to 5 (highest). The y-axis stands for the percentage of successful matches with RMS error less than 8 pixels. All degradations have 100% success rate in RetinaMatch except three high-level affine deformations. (a) ASIFT. (b) Global MI. (c) RetinaMatch.

TABLE III
TARGET REGISTRATION ERROR (TRE) OF TEMPLATE MATCHING METHODS IN EXPERIMENT 2

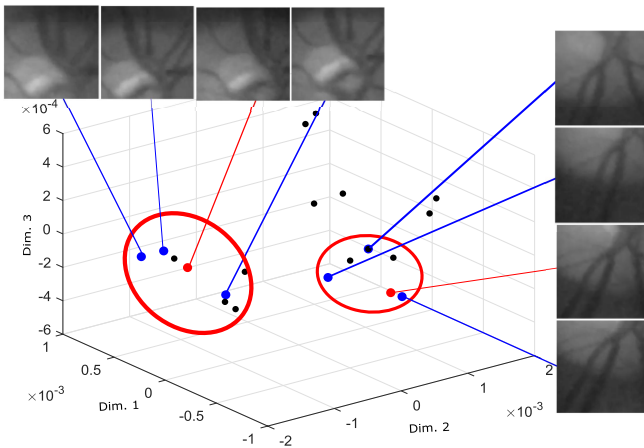|	ASIFT		MI		RetinaMatch		Observer
	Mean±SD	Success Rate	Mean±SD	Success Rate	Mean±SD	Success Rate	Variability
Without artificial features	NA	0	NA	0	3.88±1.72	94%	2.39±1.93
With artificial features	NA	0	NA	0	3.97±1.64	94%	2.39±1.93

Fig. 6. Features of images to be stitched in the top three dimensional space. Each small black dot indicates one mapped image. The colored dots in red circles show two selected samples (red) with their nearest three neighbors (blue). Note the distance is measured in the top 20 dimensional space.

from the full image independently. To obtain TRE for each image pair, we compute the root mean square of the distance between the transformed landmark points and the landmarks selected by two trained observers. The TRE results of RetinaMatch (coarse localization and final results), ASIFT and global MI are shown in Table III. Table III lists the success rate, the mean and standard deviation of TRE of successful matches and inter-observer variability. The success rate is the percentage of successful matching pairs with TRE less than 6 pixels. RetinaMatch can reach an accuracy of less than 4-pixel TRE with the observer variability, while the ASIFT and global MI cannot match the *D-eye* image successfully.

C. In Vivo D-eye Data and Mosaicked Full Image

In this experiment we match the D-eye templates onto the full image mosaicked with D-eye images. Using the stitched

panoramic image allows the use of this device at home without going to the clinic for the full fundus image as the baseline. Inhabitants of remote areas without local eye clinics having professional fundus camera facilities can benefit greatly from this technique.

1) Full Image Mosaicking: The full image in this experiment is mosaicked with 20 D-eye images using the proposed image stitching method. Based on no training for the D-eye user and other limitations of the procedure, we collected images covering the region around the optic disc. In the implementation, we used the first 20 dimensions of the features when computing the image distances in the low-dimensional space. Fig. 6 illustrates the distribution of the first three dimensions of the features. From the two examples in the figure, we can see the nearest three neighbors of the selected sample in the low dimensional space also have a large overlap in the image space. In the image patch registration, the MI-based registration method is applied. The last row of Fig. 7 shows the mosaicking result with the proposed method. The MI of the top three candidate neighbors are validated to be effective to selecting the correct neighbor. The stitched full image has a 10° FOV with the same resolution as the D-eye templates. The image blending is not our focus in this paper and the mosaicked image is not perfectly seamless.

2) Template Matching: Similar to experiment 2, we validate our method by matching 100 D-eye templates with and without pathological artificial features onto the mosaicked image. The images used for the mosaicking are not contained in the 100 template test set. The TRE results are shown in Table IV. The TRE of successful matches is less than 8 pixels. RetinaMatch can match 96% of image pairs without artifacts and 94% of image pairs with artifacts. The TRE results of success matches were not much different from the observer variability. On the other hand, ASIFT cannot find the alignment position since the detected feature points are not sufficient for matching. The MI approach has a low rate

TABLE IV

TARGET REGISTRATION ERROR (TRE) OF TEMPLATE MATCHING METHODS IN EXPERIMENT 3. THE SUCCESS RATE OF ASIFT IS 0 BECAUSE THE TEMPLATE CONTAIN AN INADEQUATE NUMBER OF FEATURES. THE SUCCESS RATE OF MI IS ALSO LOW BECAUSE THE OPTIMIZATION GETS STUCK IN LOCAL MINIMA. IN CONTRAST, RETINAMATCH HAS A CONSISTENTLY HIGH SUCCESS RATE

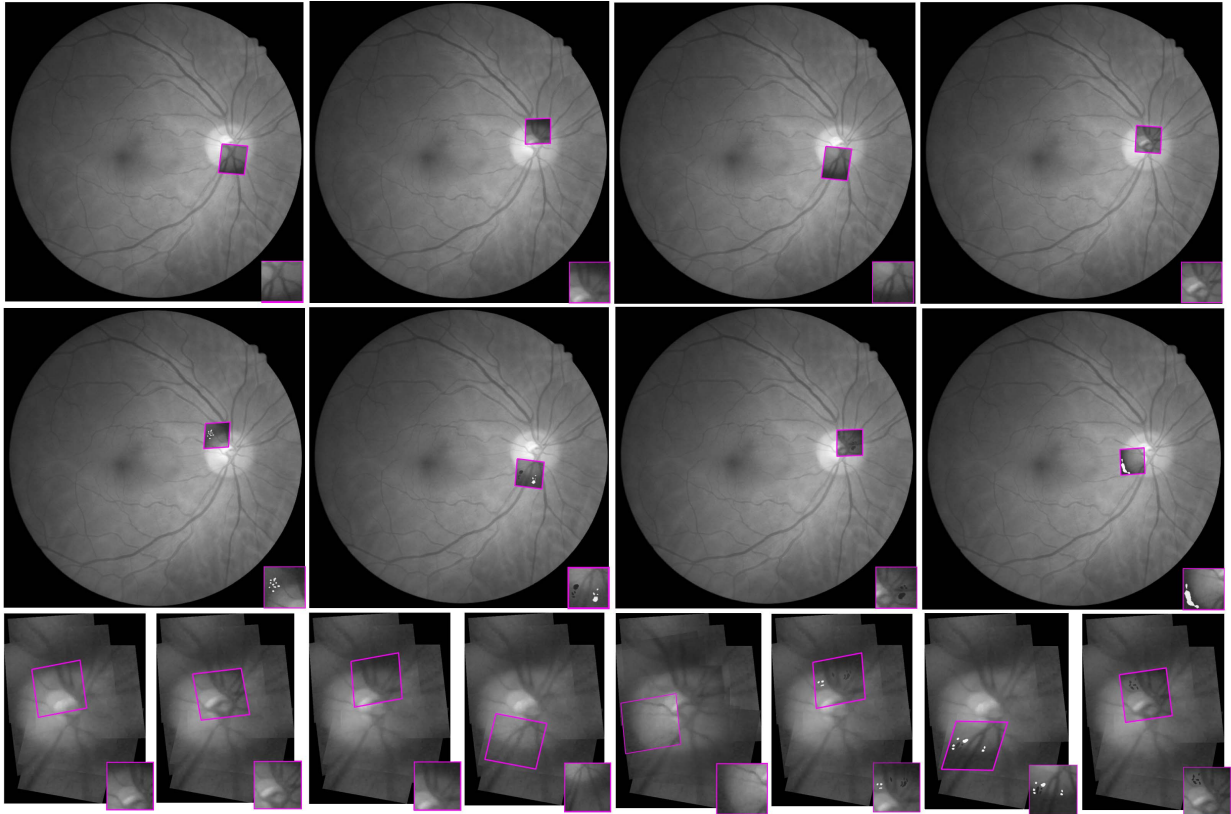|	ASIFT		MI		RetinaMatch		Observer
	Mean$\pm$SD	Success Rate	Mean$\pm$SD	Success Rate	Mean$\pm$SD	Success Rate	Variability
Without artificial features	NA	0	2.77$\pm$1.54	15%	3.06$\pm$1.65	96%	2.80$\pm$1.02
With artificial features	NA	0	3.34$\pm$1.52	8%	3.24$\pm$1.75	94%	2.80$\pm$1.02

Fig. 7. Examples of RetinaMatch results with and without artifacts. The first two rows are results of experiment 2 and the third row is experiment 3. The mapped template on the full image covers the original area and is boxed with magenta lines.

of success as well, which has a high probability to cause mis-detection of emerging changes. Fig. 7 shows examples of RetinaMatch matching results with and without artifacts in the second and third experiments.

V. CONCLUSION AND DISCUSSION

We present a new template matching method, RetinaMatch, which can be used in remote retina health monitoring with affordable imaging devices. A PCA-based coarse localization method is proposed to provide a good initialization for the MI-based registration in the template matching. In this way, RetinaMatch can obtain an accurate affine transformation between the image pair with poor quality and large size difference. As demonstrated in the simulation experiment, RetinaMatch does not handle templates with large affine deformations, with the success rate decreasing at level 4 and 5 in Fig. 5. Importantly, the template image captured by adapter-based optics with general operation will not have a

linear deformation exceeding the RetinaMatch limit, therefore we can ignore the poor performance over the third-level affine degradation. To our knowledge, this is the first report addressing template matching in retina images whose template contains unconstrained small retinal areas rather than a specific object. Further algorithm testing is needed on the smartphone or other low cost fundus imaging platforms as all current testing has been limited to a PC workstation.

To validate RetinaMatch, experiments using both human datasets with simulation and *in vivo* retina images from a case study were performed. Experiments with simulated datasets allowed evaluation of the accuracy and robustness of RetinaMatch to different levels and sequences of degradations. The *in vivo* case study ensures that our method can be applied using a consumer product. It was observed that RetinaMatch provided superior performance under different image conditions over standard ASIFT and MI algorithms. The parameters, such as the offset $f = 10$, $f' = 5$ and the dimension $d = 20$,

are independent of various datasets in the implementation of RetinaMatch, which makes it easier to translate our method to other similar imaging device besides *D-eye*.

The evaluation of the RetinaMatch accuracy is difficult in experiment 2 and 3 without ground truth. Since the goal of template matching is providing accurate alignment for downstream analysis, we use TRE as the metric. Compared with entropy based measures or the similarity measure itself, TRE measures the result intuitively in pixels and is independent of different regularization methods.

The remote monitoring of retina health with template matching is the first medical application to be proposed with RetinaMatch. Tele-ophthalmology is a promising application since many diseases are manifested at an early stage that are detectable with optical imaging of the retina. Because early stage retinal diseases do not present with symptoms, routine screening is important for early detection, which requires both high sensitivity and even higher specificity. The adapter-based optics and the digital cameras from smart phones provide an efficient and economic approach to capture retina images regularly at home. The images of the current state can be mapped with RetinaMatch and then compared with the previous state. With regular screening, the process of lesion formation and therapeutic treatment can be monitored over time. In the experiment, *D-eye* is chosen just as one low-cost example among many others with small FOV on undilated pupils [5]. Similar fundus imaging techniques can also be implemented in emerging commercial VR, AR, and mixed reality headsets that will be widespread in the future.

There are different kinds of retina lesions that can be screened with portable fundus cameras. In the medical example of monitoring hypertension, the larger arteries constrict and the venous vessels enlarge in diameter [4]. For example, the larger blood vessel cross-sectional diameter is about 20 pixels in the case study, and a change with hypertension will be in the range of 10-60%, so we are looking for over 2-pixel changes from baseline over time. In Tables III and IV, the TRE is shown to be extremely low and most errors are below 2 pixels (1.8 arcmins) excluding observer variability. With advanced trend analytics [45], we can expect template matching errors to be well below a threshold of clinical significance. For more precise vessel width measurement, RetinaMatch can be combined with vessel segmentation, as described in our previous publication [46]. The vessels of interest can be located on the current templates and the corresponding vessel width is then obtained by segmentation around the mapped location. Note the vessel segmentation here is applied on very small retina patches (20×20 pixels), which is more robust and accurate than segmentation of wide FOV retina images. The segmentation error in [46] is less than 1 pixel, which has been presented using *D-eye* images. Xu *et al.* [47] proposed the vessel width segmentation and measurement on retina imaging acquired from the low quality fundus camera as well. They also report similar 1-pixel accuracy. However, the imaging device they used produced higher quality retinal images, having five times larger FOV than the *D-eye*. The biomarkers of abusive head trauma (AHT) is another example. The most common retinal manifestation of AHT is multiple retinal hemorrhages in multiple layers of the retina [48]. Matching the captured images onto the full retina image, The hemorrhagic spots can be easily segmented after the subtraction of the current retina regions and previous status. The AHT then can be recognized automatically when such spots are detected with portable fundus cameras.

RetinaMatch may be used in other medical image applications for template matching. For example, in the case of endoscopic guidance of therapy by a surgical robot [49], the current limited-sized FOV can be matched onto the panorama for endoscope localization. Thus, this image template matching technique can be used to create a more reliable closed-loop control for the robot arm and surgical tool guidance.

REFERENCES

[1] S. Wang, K. Jin, H. Lu, C. Cheng, J. Ye, and D. Qian, "Human visual system-based fundus image quality assessment of portable fundus camera photographs," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1046–1055, Apr. 2016.

[2] L. Shi, H. Wu, J. Dong, K. Jiang, X. Lu, and J. Shi, "Telemedicine for detecting diabetic retinopathy: A systematic review and meta-analysis," *Brit. J. Ophthalmology*, vol. 99, no. 6, pp. 823–831, Jun. 2015.

[3] I. N. Figueiredo, S. Kumar, C. M. Oliveira, J. D. Ramos, and B. Engquist, "Automated lesion detectors in retinal fundus images," *Comput. Biol. Med.*, vol. 66, pp. 47–65, Nov. 2015.

[4] R. Kawasaki *et al.*, "Retinal vessel diameters and risk of hypertension: The multiethnic study of atherosclerosis.," *J. Hypertension*, vol. 27, no. 12, pp. 93–2386, Dec. 2009.

[5] N. Panwar *et al.*, "Fundus photography in the 21st century—A review of recent technological advances and their implications for worldwide healthcare," *Telemed. e-Health*, vol. 22, no. 3, pp. 198–208, Mar. 2016.

[6] W. Fink, M. A. Tarbell, and K. Garcia, "Smart ophthalmics: The future in tele-ophthalmology has arrived," *Proc. SPIE*, vol. 9836, May 2016, Art. no. 98360W.

[7] K. Roesch, T. Swedish, and R. Raskar, "Automated retinal imaging and trend analysis—A tool for health monitoring," *Clin. Ophthalmology*, vol. 11, pp. 1015–1020, May 2017.

[8] C. A. Ludwig, "The future of automated mobile eye diagnosis," Byers Eye Inst., Stanford Univ. School Med., Palo Alto, CA, USA, Tech. Rep. Journal MTM 5:2:44-50, 2016. doi: 10.7309/jmtm.5.2.7.

[9] H. Yu *et al.*, "Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 644–657, Jul. 2012.

[10] X. Zhang *et al.*, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, Oct. 2014.

[11] A. D. Mora, J. Soares, and J. M. Fonseca, "A template matching technique for artifacts detection in retinal images," in *Proc. 8th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2013, pp. 717–722.

[12] C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1379–1394, Nov. 2003.

[13] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1531–1546, Dec. 2006.

[14] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. C. T. Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognit.*, vol. 40, no. 7, pp. 2063–2076, Jul. 2007.

[15] I. N. Figueiredo *et al.*, "Automated retina identification based on multiscale elastic registration," *Comput. Biol. Med.*, vol. 79, pp. 130–143, Dec. 2016.

[16] I. N. Figueiredo, J. S. Neves, S. Moura, C. M. Oliveira, and J. D. Ramos, "Pattern classes in retinal fundus images based on function norms," in *Proc. Int. Symp. Comput. Modeling Objects Represented Images*. Cham, Switzerland: Springer, 2014, pp. 95–105.

[17] Y. Wang, J. Shen, W. Liao, and L. Zhou, "Automatic fundus images mosaic based on SIFT feature," in *Proc. 3rd Int. Congr. Image Signal Process. (CISP)*, vol. 6, Oct. 2010, pp. 2747–2751.

[18] C. L. Tsai, C.-Y. Li, G. Yang, and K.-S. Lin, "The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 636–649, Mar. 2010.

[19] G. Wang, Z. Wang, Y. Chen, and W. Zhao, "Robust point matching method for multimodal retinal image registration," *Biomed. Signal Process. Control*, vol. 19, pp. 68–76, May 2015.

[20] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration based on keypoint correspondences, spherical eye modeling and camera pose estimation," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 5650–5654.

[21] K. J. Friston, J. Ashburner, C. D. Frith, J.-B. Poline, J. D. Heather, and R. S. J. Frackowiak, "Spatial registration and normalization of images," *Hum. Brain Mapping*, vol. 3, no. 3, pp. 165–189, Mar. 1995.

[22] A. V. Cideciyan, "Registration of ocular fundus images: An algorithm using cross-correlation of triple invariant image descriptors," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 1, pp. 52–58, Jan./Feb. 1995.

[23] Y.-M. Zhu, "Mutual information-based registration of temporal and stereo retinal images using constrained optimization," *Comput. Methods Programs Biomed.*, vol. 86, no. 3, pp. 210–215, Jun. 2007.

[24] E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, "Randomized algorithms for the low-rank approximation of matrices," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 51, pp. 20167–20172, Dec. 2007.

[25] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011.

[26] N. B. Erichson, A. Mendible, S. Wihlborn, and J. N. Kutz, "Randomized nonnegative matrix factorization," *Pattern Recognit. Lett.*, vol. 104, pp. 1–7, Mar. 2018.

[27] *Structured Analysis of the Retina*. Accessed: May 15, 2018. [Online]. Available: http://www.ces.clemson.edu/~ahoover/stare/

[28] F. R. S. K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Philosoph. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[29] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, p. 417, 1933.

[30] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, vol. 149. Philadelphia, PA, USA: SIAM, 2016.

[31] N. B. Erichson, S. Voronin, S. L. Brunton, and J. N. Kutz, "Randomized matrix decompositions using R," 2016, *arXiv:1608.02148*. [Online]. Available: https://arxiv.org/abs/1608.02148

[32] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11–37, May 2011.

[33] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust PCA in image and video processing," *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, Aug. 2018.

[34] A. Baghaie, R. M. D'souza, and Z. Yu, "Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal OCT images," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 226–230.

[35] Y. Fu, C. Wang, Y. Wang, B. Chen, Q. Peng, and L. Wang, "Automatic detection of longitudinal changes for retinal fundus images based on low-rank decomposition," *J. Med. Imag. Health Informat.*, vol. 8, no. 2, pp. 284–294, Feb. 2018.

[36] P. Thevenaz and M. A. Unser, "Spline pyramids for intermodal image registration using mutual information," *Proc. SPIE*, vol. 3169, Oct. 1997, pp. 236–248.

[37] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imag.*, vol. 22, no. 1, pp. 120–128, Jan. 2003.

[38] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, 2009.

[39] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw.* New York, NY, USA: Springer, Oct. 1997, pp. 583–588.

[40] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[41] L. J. Van Rijn, "Torsional eye movements in humans," *Investigational New Drugs-Invest New Drug*, Jan. 1994.

[42] T. Walter, J.-C. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE Trans. Med. Image*, vol. 21, no. 10, pp. 1236–1243, Oct. 2002.

[43] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[44] J. M. Fitzpatrick, J. B. West, and C. R. Maurer, Jr., "Predicting error in rigid-body point-based registration," *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 694–702, Oct. 1998.

[45] X. Dai and M. Bikdash, "Trend analysis of fragmented time series for mhealth apps: Hypothesis testing based adaptive spline filtering method with importance weighting," *IEEE Access*, vol. 5, pp. 27767–27776, 2017.

[46] C. Gong, J. P. Kelly, L. Trutoiu, B. Schowengerdt, S. Brunton, and E. Seibel, "Measurement of retinal vessel width in tele-ophthalmology for mobile health monitoring," *Investigative Ophthalmology Vis. Sci.*, vol. 59, no. 9, p. 4624, Jul. 2018.

[47] X. Xu *et al.*, "Smartphone-based accurate analysis of retinal vasculature towards point-of-care diagnostics," *Sci. Rep.*, vol. 6, Oct. 2016, Art. no. 34603.

[48] I. H. Yusuf, J. K. Barnes, T. H. Fung, J. S. Elston, and C. K. Patel, "Non-contact ultra-widefield retinal imaging of infants with suspected abusive head trauma," *Eye*, vol. 31, no. 3, p. 353, Mar. 2017.

[49] D. Hu, Y. Gong, E. J. Seibel, L. N. Sekhar, and B. Hannaford, "Semi-autonomous image-guided brain tumour resection using an integrated robotic system: A bench-top study," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 1, p. e1872, Feb. 2018.