

Combining Feature Correspondence with Parametric Chamfer Alignment: Hybrid Two-Stage Registration for Ultra-Widefield Retinal Images

Li Ding ¹, Tony Kang ¹, Ajay E. Kuriyan ¹, Rajeev S. Ramchandran ¹, Charles C. Wykoff ¹, and Gaurav Sharma ¹

¹Affiliation not available

October 30, 2023

Abstract

We propose a novel hybrid framework for registering retinal images in the presence of extreme geometric distortions that are commonly encountered in ultra-widefield (UWF) fluorescein angiography. Our approach consists of two stages: a feature-based global registration and a vessel-based local refinement. For the global registration, we introduce a modified RANSAC algorithm that jointly identifies robust matches between feature keypoints in reference and target images and estimates a polynomial geometric transformation consistent with the identified correspondences. Our RANSAC modification particularly improves feature point matching and the registration in peripheral regions that are most severely impacted by the geometric distortions. The second local refinement stage is formulated in our framework as a parametric chamfer alignment for vessel maps obtained using a deep neural network. Because the complete vessel maps contribute to the chamfer alignment, this approach not only improves registration accuracy but also aligns with clinical practice, where vessels are typically a key focus of examinations. We validate the effectiveness of the proposed framework on a new UWF fluorescein angiography (FA) dataset and on the existing narrow-field FIRE (fundus image registration) dataset and demonstrate that it significantly outperforms prior retinal image registration methods. The proposed approach enhances the utility of large sets of longitudinal UWF images by enabling: (a) automatic computation of vessel change metrics and (b) standardized and co-registered examination that can better highlight changes of clinical interest to physicians.

Combining Feature Correspondence with Parametric Chamfer Alignment: Hybrid Two-Stage Registration for Ultra-Widefield Retinal Images

Li Ding, Tony D. Kang, Ajay E. Kuriyan, Rajeev S. Ramchandran, Charles C. Wykoff, and Gaurav Sharma

Abstract—We propose a novel hybrid framework for registering retinal images in the presence of extreme geometric distortions that are commonly encountered in ultra-widefield (UWF) fluorescein angiography. Our approach consists of two stages: a feature-based global registration and a vessel-based local refinement. For the global registration, we introduce a modified RANSAC algorithm that jointly identifies robust matches between feature keypoints in reference and target images and estimates a polynomial geometric transformation consistent with the identified correspondences. Our RANSAC modification particularly improves feature point matching and the registration in peripheral regions that are most severely impacted by the geometric distortions. The second local refinement stage is formulated in our framework as a parametric chamfer alignment for vessel maps obtained using a deep neural network. Because the complete vessel maps contribute to the chamfer alignment, this approach not only improves registration accuracy but also aligns with clinical practice, where vessels are typically a key focus of examinations. We validate the effectiveness of the proposed framework on a new UWF fluorescein angiography (FA) dataset and on the existing narrow-field FIRE (fundus image registration) dataset and demonstrate that it significantly outperforms prior retinal image registration methods. The proposed approach enhances the utility of large sets of longitudinal UWF images by enabling: (a) automatic computation of vessel change metrics and (b) standardized and co-registered examination that can better highlight changes of clinical interest to physicians.

Index Terms—Image registration, vessel detection, fluorescein angiography, retinal image analysis, RANSAC

I. INTRODUCTION

Retinal image registration is one of the crucial tasks in ophthalmological image analysis that facilitates several clinical applications. Registration of multiple narrow-field color fundus

images captured during a clinical visit has been used to produce a montage with large field-of-view (FOV) [1]. Registration of retinal images captured over a series of longitudinal visits has also been used for quantitative and qualitative assessment of temporal changes in vasculature [2]–[4].

The goal of automated retinal image registration is to determine a geometric transformation between the spatial coordinates of images captured from different viewpoints so as to align their content. The problem has been studied extensively and existing works can be broadly grouped into three categories: intensity-based, keypoint-based, and segmentation-based. Intensity-based approaches estimate the registration transform by maximizing a similarity metric between the image pairs under the registration transform. Various metrics have been proposed, including mutual information [5], [6], mean squared difference [7], cross-correlation [8], and phase correlation [9]. Keypoint-based methods estimate the registration transform from correspondences of feature keypoints extracted from the retinal images, where RANSAC [10] is typically used to establish correspondences robustly. Both hand-crafted feature detectors and deep-learning based feature detectors have been used for retinal image registration [11]–[14]. Methods in the third category, detect binary vessel maps from the retinal images, which are then used to estimate the registration transform via a number of alternative methodologies. Specifically, the problem has been formulated as deformable line-based registration for skeletonized vessel maps [15], as two-dimensional registration of vessel pixel point clouds [16], and as topological matching of the vessel connectivity structure represented as a graph [17], [18]. In addition to 2D transformations between image coordinates, a few methods have also formulated the registration problem as joint estimation of camera poses and 3D retinal structure [19], [20]. An iterative framework has also been recently proposed [21], which leverages an available deep-learning based vessel detector for one imaging modality to jointly perform cross-modal registration and weakly-supervised vessel detection for another retinal imaging modality.

Most of the prior works are designed for narrow-field retinal images, which constitute the predominant type of retinal imagery. With advances in imaging techniques, ultra-widefield (UWF) retinal images, such as UWF fluorescein angiography (FA) and UWF fundus photography (FP), are also commonly being utilized for clinical assessment [22]–[24]. Compared with narrow-field retinal images that provide a FOV of 30° to 50°, UWF images capture a substantially larger FOV (up to

The work was supported in part by a University of Rochester Research Award, by a distinguished researcher award from the New York state funded Rochester Center of Excellence in Data Science (CoE #3B C160189) at the University of Rochester, by an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, and grant P30EY001319-35 from the National Institutes of Health.

L. Ding and G. Sharma are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA (e-mail: {l.ding, gaurav.sharma}@rochester.edu).

T. D. Kang and R. S. Ramchandran are with the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: {tony_kang, rajeev_ramchandran}@urmc.rochester.edu).

A. E. Kuriyan is with the Mid Atlantic Retina, Retina Service of Wills Eye Hospital, Thomas Jefferson University, Philadelphia, PA 19107 & the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: ajay.kuriyan@gmail.com).

C. C. Wykoff is with Retina Consultants of Texas and Blanton Eye Institute, Houston Methodist Hospital & Weill Cornell Medical College, Houston, TX 77030, USA (e-mail: charleswykoff@gmail.com).

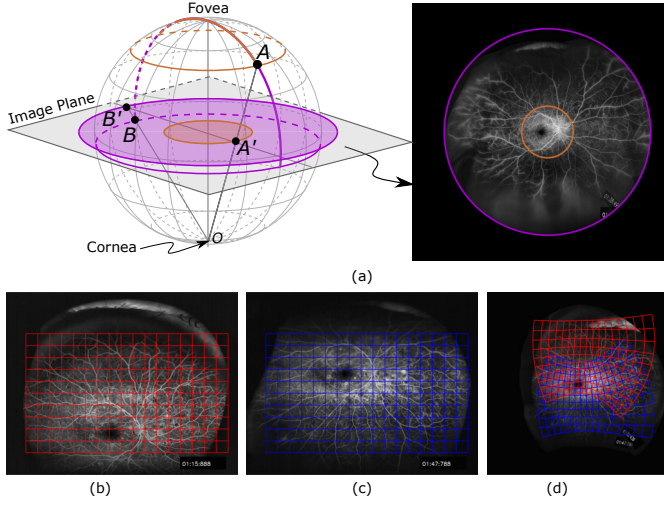


Fig. 1: Stereographic projection and extreme geometric distortion between individual UWF images and montages. (a) Stereographic projection for UWF (purple) and narrow field (brown) retinal images. Points A and B on the retinal surface project to A' and B' on the image plane, respectively; (b) and (c) two individual UWF FA retinal images; (d) the corresponding stereographic projected montage FA image. The blue and red grids in (d) correspond to the grids in (b) and (c), respectively, and illustrate the extreme geometric distortions that the registration must handle.

200°), typically covering the retinal surface from the posterior pole to the periphery in a single image [25]. Significant variation and geometric distortion depending on the direction of capture are inevitably encountered for the larger FOV. While the geometric distortions cannot be completely eliminated in mapping from the nearly spherical retina to a planar image, a standardized representation stereographic projection [26], [27] has been developed that also combines information across multiple images in a single montage. Figure 1 illustrates the stereographic projection process and highlights the extreme geometric distortions seen in individual capture UWF images. Because generating montage images has challenges and can introduce artifacts, currently the clinical analysis relies more on the original UWF images. However, individual UWF images are numerous and are not well aligned due to patient eye movement, which limits the physician’s ability to examine multiple UWF images in detail. Therefore, automated registration of UWF image to the montage representation is desirable to standardize the geometry and to produce consistent and repeatable metric quantification.

In this paper, we address the problem of registering UWF retinal images with a stereographic projected montage image and make the following contributions. We propose a novel two-stage hybrid registration approach for UWF retinal images that uses the intensity image and the corresponding binary vessel segmentation maps for the first and second stages, respectively. The first stage coarsely aligns the grayscale UWF images using feature-matching based global registration, and the second stage provides fine registration using vessel-based local chamfer alignment. The methodology enables precise

registration of UWF images despite the strong changes in imaging geometries typical for such images. The choice of vessel based chamfer alignment for the second stage makes is also particularly well matched with clinical scenarios requiring longitudinal comparisons between vessel structures. We propose a modification to the well-known RANSAC algorithm that separates sample and consensus sets to provide better estimates of the global transformation parameters for our first stage, which significantly improves accuracy in the peripheral regions. To evaluate retinal image registration accuracy and to facilitate further research, we introduce a new dataset of UWF FA retinal images called FLoRI21, which presents more realistic challenges for image registration relevant in clinical applications. We present an extensive evaluation of the proposed method on the new FLoRI21 dataset and on the existing narrow-field FIRE dataset [28] and, for both datasets, demonstrate significant improvement over prior methods.

The rest of this paper is organized as follows. In Section II, we present the proposed hybrid retinal registration method. In Section III, we conduct a series of analysis of the proposed method and compare with existing works on retinal image registration. Section IV summarizes the concluding remarks.

II. HYBRID REGISTRATION APPROACH

The proposed hybrid retinal image registration technique, which is illustrated in Fig. 2, leverages the complementary strengths of the feature-based global registration and the vessel-based local registration. Used in conjunction with robust techniques, such as RANSAC, feature-based global registration avoids local minima for transformation parameter estimates. However, a global transformation cannot handle the extreme geometric distortion that is introduced from mapping the approximately spherical retinal surface to the image plane. Thus, invariably, the global transformation does not provide adequate registration accuracy. By limiting the scope to a restricted smaller spatial region, the local registration, however, can model the residual mismatch remaining after the global registration. Specifically, for the local registration, we utilize the anatomical segmentation of retinal vessels detected from input images and partition the coarsely aligned vessel pairs into overlapping patches. Unlike the global registration that relies on a set of sparse feature correspondences, the local registration step optimizes a dense vessel pixel agreement metric for each local patch via the parametric chamfer alignment.

Next, we describe details for the proposed feature-based global registration and the vessel-based local registration.

A. Global Feature-based Registration

The global registration step aims to estimate a transformation \tilde{T}_β between a reference image I_r (fixed geometry) and a target image I_t . Since the imaging geometry of the reference and the target images are different and unavailable, we adopt a parametric polynomial transformation that is flexible enough to capture non-rigid changes in geometry. In this step, feature points are first extracted from the reference image and the target image, shown as the cyan and the magenta points in the global registration block in Fig. 2, respectively. We

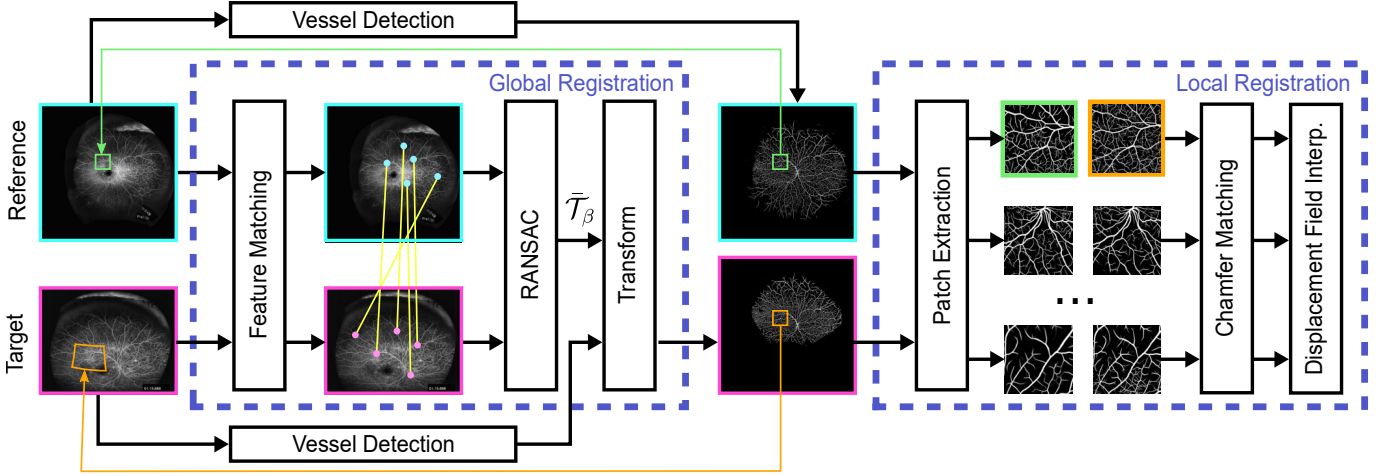


Fig. 2: Overview of the proposed hybrid registration method. Best viewed in color.

denote the set of the detected reference feature points by $\mathcal{F}_r = \{(\mathbf{f}_i^{(r)}, \mathbf{p}_i^{(r)})\}_{i=1}^{N_r}$, where $\mathbf{f}_i^{(r)} \in \mathbb{R}^d$ is the d -dimensional feature descriptor, $\mathbf{p}_i^{(r)} \in \mathbb{R}^2$ is the location of feature point, and N_r is the number of detected reference feature points. Similarly, the set of target feature points is represented as $\mathcal{F}_t = \{(\mathbf{f}_j^{(t)}, \mathbf{p}_j^{(t)})\}_{j=1}^{N_t}$. Alternative feature descriptors can be used in the proposed method, including those that are hand-crafted (e.g. SIFT [29]) or learned (e.g. SuperPoints [30]).

From the detected features \mathcal{F}_r and \mathcal{F}_t , we establish two sets of (tentative) corresponding points. The first set \mathcal{S}_c is the consensus set containing the nearest feature matches and the second set \mathcal{S}_s is the sample set in which only distinctive features are included. Specifically, for each reference feature descriptor $\mathbf{f}_i^{(r)}$, we first search its nearest neighbor $\mathbf{f}_{j_{i,1}}^{(t)}$ in the feature space from the target feature set \mathcal{F}_t , and add the corresponding location pair $(\mathbf{p}_i^{(r)}, \mathbf{p}_{j_{i,1}}^{(t)})$ to the consensus set \mathcal{S}_c . Next, if $\mathbf{f}_{j_{i,1}}^{(t)}$ is close to $\mathbf{f}_i^{(r)}$ by a significant margin compared to other features in \mathcal{F}_t , i.e., if

$$\frac{\|\mathbf{f}_i^{(r)} - \mathbf{f}_{j_{i,1}}^{(t)}\|}{\|\mathbf{f}_i^{(r)} - \mathbf{f}_j^{(t)}\|} < \tau_1, \quad (1)$$

for all $j \neq j_{i,1}$ for a suitably chosen threshold $\tau_1 < 1$, we declare it a *distinctive match* and add it to the sample set \mathcal{S}_s . To make the feature matching robust, we perform the bidirectional feature matching by swapping the reference features and the target features.

The consensus and the sample sets have different characteristics. While the mismatch rate in the sample set \mathcal{S}_s is relatively low, the number of matched feature points is small, especially in the peripheral regions in which the registration is prone to error. On the other hand, the consensus set \mathcal{S}_c includes a large number of matched features, even though the mismatch rate is high as well. Using these two sets of tentative matched features, we propose a modification to the RANSAC algorithm to estimate parameters β for a polynomial registration transformation. The same overall methodology applies for different order polynomial transformations, though

the number of parameters and the form of the transformation depends on the specific choice.

In the proposed modified RANSAC, we first randomly select N_β matched points from the sample set \mathcal{S}_s . We determine the optimal transformation parameter β such that the transformed target coordinates closely approximate the corresponding reference points, i.e., $\mathbf{p}_n^{(r)} \approx \bar{\mathcal{T}}_\beta(\mathbf{p}_n^{(t)})$ for $n = 1, 2, \dots, N_\beta$. The least squares solution is

$$\beta = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}. \quad (2)$$

Detailed expressions for \mathbf{A} and \mathbf{b} are provided in Section S.II in the Supplementary Material for the second, third, and fourth order polynomial transformations that we used in our experiments.

We then determine the inlier points from the consensus set \mathcal{S}_c that lie within a distance threshold τ_2 under the estimated transformation parameters. The random selection is repeated for N_{trial} times and the selection with the largest number of inliers is used to determine the optimal global transformation parameters. The overall algorithm for the global transformation estimation is summarized in Algorithm 1. We note that, although simple, the proposed modification to the conventional RANSAC algorithm using separate sample and consensus sets provides significant gains in practice (as we demonstrate in our experimental results).

Once the global transformation parameters have been estimated, we apply the transformation $\bar{\mathcal{T}}_\beta$ to the target vessel map \mathbf{V}_t to obtain the warped version \mathbf{V}_{t_0} . The resulting coarsely aligned vessel pair $(\mathbf{V}_r, \mathbf{V}_{t_0})$ is then fed to the local registration step.

B. Local Vessel-Based Parametric Chamfer Alignment

As indicated in Section I, a global transformation is usually insufficient to model the extreme geometric distortion in the UWF images. Therefore, in the second step, we refine the registration at the local patch level via parametric chamfer alignment of retinal vessels.

We first partition the coarsely aligned vessel maps \mathbf{V}_r and \mathbf{V}_{t_0} into K pairs of overlapping patches (see the local

Algorithm 1: Global transformation estimation

Given : Feature matching threshold τ_1 , inlier distance threshold τ_2 , number of RANSAC trials N_{trial} , and number of pairs N_β to estimate transformation parameters β

Input : Reference image \mathbf{I}_r , target image \mathbf{I}_t

Output: Parameters β for optimal global registration transformation

```

1  $N_{\text{best}} \leftarrow 0$ ,  $\mathcal{S}_c \leftarrow \emptyset$ ,  $\mathcal{S}_s \leftarrow \emptyset$ ;
2 Extract feature descriptors  $\mathcal{F}_r$  from  $\mathbf{I}_r$  and  $\mathcal{F}_t$  from  $\mathbf{I}_t$ ;
  // Match features from  $\mathcal{F}_r$  to  $\mathcal{F}_t$ 
3  $(\mathcal{S}_c, \mathcal{S}_s) \leftarrow (\mathcal{S}_c, \mathcal{S}_s) \cup \text{FMatches}(\mathcal{F}_r, \mathcal{F}_t, \tau_1, 0)$ ;
  // Match features from  $\mathcal{F}_t$  to  $\mathcal{F}_r$ 
4  $(\mathcal{S}_c, \mathcal{S}_s) \leftarrow (\mathcal{S}_c, \mathcal{S}_s) \cup \text{FMatches}(\mathcal{F}_t, \mathcal{F}_r, \tau_1, 1)$ ;
  // Transformation Estimation via RANSAC
5 for  $n = 1 : N_{\text{trial}}$  do
6   Randomly sample  $N_\beta$  pairs from  $\mathcal{S}_s$  to estimate
     transformation parameters  $\beta_n$  using (2);
     // Find inlier matches from the consensus set
7    $\mathcal{S}_n \leftarrow \{(\mathbf{p}^{(r)}, \mathbf{p}^{(t)}) \in \mathcal{S}_c \mid \|\mathbf{p}^{(r)} - \tilde{\mathcal{T}}_{\beta_n}(\mathbf{p}^{(t)})\| < \tau_2\}$ ;
     // Maintain inliers corresponding to the best fit
8   if  $|\mathcal{S}_n| > N_{\text{best}}$  then
9      $N_{\text{best}} \leftarrow |\mathcal{S}_n|$ ,  $\mathcal{S}_{\text{inlier}} \leftarrow \mathcal{S}_n$ ;
10  end
11 end
12 Estimate global transformation parameters  $\beta$  from
     $\mathcal{S}_{\text{inlier}}$  using (2);
    // Unidirectional feature matching function
13 Function  $\text{FMatches}(\mathcal{F}_1, \mathcal{F}_2, \tau, \Pi)$ 
14   foreach  $(\mathbf{f}_i^{(1)}, \mathbf{p}_i^{(1)}) \in \mathcal{F}_1$  do
15     Find the nearest feature neighbor  $\mathbf{f}_{j_{i,1}}^{(2)}$  and the
        second nearest feature neighbor  $\mathbf{f}_{j_{i,2}}^{(2)}$  from  $\mathcal{F}_2$ ;
        // Add nearest neighbor matches to consensus set
16      $\mathcal{S}_c \leftarrow \mathcal{S}_c \cup \{\text{Swap}(\mathbf{p}_1^i, \mathbf{p}_2^{j_{i,1}}, \Pi)\}$ ;
17     if  $\frac{\|\mathbf{f}_i^{(1)} - \mathbf{f}_{j_{i,1}}^{(2)}\|}{\|\mathbf{f}_i^{(1)} - \mathbf{f}_{j_{i,2}}^{(2)}\|} < \tau$  then
18       // Add distinctive matches to sample set
19        $\mathcal{S}_s \leftarrow \mathcal{S}_s \cup \{\text{Swap}(\mathbf{p}_i^{(1)}, \mathbf{p}_{j_{i,1}}^{(2)}, \Pi)\}$ 
20     end
21   end
22   return  $(\mathcal{S}_c, \mathcal{S}_s)$ ;
23 end
24 Function  $\text{Swap}(a, b, \Pi)$  // Conditional swap on  $\Pi = 1$ 
25   if  $\Pi = 1$  then return  $(b, a)$  else return  $(a, b)$ ;
26 end
  
```

registration block in Fig. 2). The number K is determined based on the patch size and the stride size, i.e., $K = (H - P + S)(W - P + S)/S^2$, where H , W , P , and S are image height, image width, patch size, and stride size, respectively. The parametric chamfer alignment is performed for each pair of vessel maps separately. For each patch k , let $\mathcal{Q}_{r,k} = \{\mathbf{p}_i^{(r,k)}\}_{i=1}^{M_{r,k}}$ be the set of vessel pixel locations in the reference patch where $\mathbf{p}_i^{(r,k)}$ is the pixel location in the local patch coordinate system and $M_{r,k}$ is the number of vessel

pixels. Similarly, we have a set of vessel pixel locations in the coarsely aligned target vessel patch $\mathcal{Q}_{t,k} = \{\mathbf{p}_j^{(t,k)}\}_{j=1}^{M_{t,k}}$

Conventional chamfer alignment [31] estimates the transformation parameters β_k that minimizes the mean-squared chamfer distance from the target vessel pixels in $\mathcal{Q}_{t,k}$ to the reference vessel pixels in $\mathcal{Q}_{r,k}$. For each target vessel pixel j , the squared chamfer distance $d_j(\beta_k)$ is calculated as the minimum squared distance between its transformed location $\tilde{\mathcal{T}}_{\beta_k}(\mathbf{p}_j^{(t,k)})$ and the nearest pixel location in $\mathcal{Q}_{r,k}$. This formulation, however, is not robust to outlier vessel pixels that only exist in $\mathcal{Q}_{t,k}$ and do not have correspondence in $\mathcal{Q}_{r,k}$. To handle the outlier pixels, we adopt a probabilistic framework for the parametric chamfer alignment [32]. Specifically, we further associate each target vessel pixel j with a latent variable $z_j^k \in \{0, 1\}$ to indicate if the pixel j has a correspondence in the reference vessel set ($z_j^k = 1$) or not ($z_j^k = 0$) and refer to the pixels in the two groups as *inliers* and *outliers*, respectively. We model the prior probability of $p(z_j^k = 1)$ as a Bernoulli distribution with unknown parameter π_k . Moreover, we assume that the registration error $d_j(\beta_k)$ for inlier vessel points should be close to 0 and thus model the conditional probability $p(d_j | z_j^k = 1)$ as an exponential distribution with unknown parameter λ_k . For outlier vessel points, the registration error $p(d_j | z_j^k = 0)$ is modeled as a uniform distribution over $[0, P^2]$.

We apply the Expectation Maximization (EM) algorithm [33] to obtain maximum-likelihood estimates of the unknown parameters π_k , λ_k , and β_k . The EM algorithm iterates between two steps: the Expectation step (E-step) and the Maximization step (M-step). In the E-step, we calculate the posterior probabilities $p_j^k = p(z_j^k = 1 | \pi_k, \lambda_k, \beta_k, d_j)$ using Bayes rule. In the M-step, we maximize the expectation of the complete-data log likelihood and obtain the updated parameters $\pi_k = \sum_{j=1}^{M_{t,k}} p_j^k / M_{t,k}$, $\lambda_k = \sum_{j=1}^{M_{t,k}} p_j^k / \sum_{j=1}^{M_{t,k}} p_j^k d_j$, and

$$\hat{\beta}_k = \arg \min_{\beta} \frac{1}{M_{t,k}} \sum_{j=1}^{M_{t,k}} p_j^k d_j(\beta). \quad (3)$$

The minimization of (3) corresponds to a probabilistic chamfer alignment in which the registration error d_j for each pixel is weighted by the posterior probability p_j that it is an inlier pixel. Using the EM framework, the distance metric for probabilistic chamfer alignment concentrates on the inlier pixels and reduces the impact of outlier pixels, which results in robust parameter estimates. In addition, the computation of the objective function in (3) can be done efficiently using the distance transform [34]. We provide detailed derivations of the posterior probability and the parameter updates in Section S.III in the Supplementary Material.

To fuse the multiple transformation parameters β_k estimated from the overlapping patches, we propose to perform the displacement vector field interpolation, which is illustrated in Fig. 3. For each patch k , we determine a local displacement vector field ϕ_k . The local displacement vector field at pixel location in the reference patch represents the vector offset from this location to the corresponding location in the target patch.

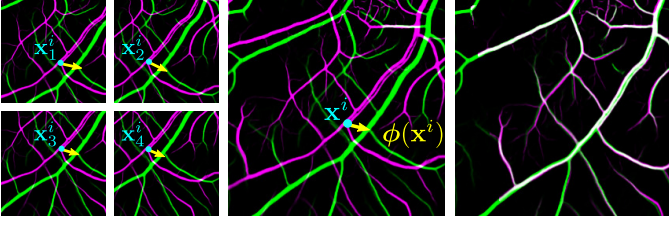


Fig. 3: Illustration of the displacement vector field interpolation. The green and the magenta pixels are the vessels in the reference and the target images, respectively. In the first two columns, the cyan point shows a target pixel in four patches and the yellow arrow is the local displacement vector estimated from each patch. The interpolated displacement vector is shown in the third column. The rightmost image shows the result from the local registration, where the white pixels are the common vessels between the image pairs.

The local displacement vector $\phi_k(\mathbf{x}_i^k)$ for the pixel location \mathbf{x}_i^k can be obtained from the estimated transformation \tilde{T}_{β_k} as

$$\phi_k(\mathbf{x}_i^k) = \tilde{T}_{\beta_k}^{-1}(\mathbf{x}_i^k) - \mathbf{x}_i^k, \quad (4)$$

where $\tilde{T}_{\beta_k}^{-1}$ is the inverse of the geometric transformation \tilde{T}_{β_k} . The interpolated displacement vector for pixel \mathbf{x}_i is obtained as

$$\phi(\mathbf{x}_i) = \frac{1}{|\mathcal{N}_i|} \sum_{k \in \mathcal{N}_i} \phi_k(\mathbf{x}_i^k), \quad (5)$$

where \mathcal{N}_i is the set of patches that contain the pixel \mathbf{x}_i . To obtain the warped target image $\mathbf{I}_{t \rightarrow r}$ that is align with the reference image \mathbf{I}_r , we transform the target image \mathbf{I}_t according to the interpolated displacement vector field ϕ .

III. EXPERIMENTAL RESULTS

In this section, we first describe the datasets used for quantitative assessment and introduce a new UWF FA dataset, FLoRI21 [35], for retinal image registration. Next, we define the metrics for quantitative evaluation and list the existing retinal image registration methods for comparison. We then conduct extensive experiments to evaluate the proposed method for retinal image registration. We first study the effectiveness of the proposed two-stage method via a series of in depth ablation studies and analysis (Section III-C) and then perform a comparative evaluation of the proposed method against existing retinal image registration methods (Section III-D).

A. Dataset

We validate the proposed method for retinal image registration on a new ultra-widefield FA dataset, FLoRI21, and on the existing narrow-field FIRE (fundus image registration) dataset [28].

The FLoRI21 (Fluorescein-angiography Longitudinal Retinal Image 2021) dataset provides 15 pairs of UWF FA images that were collected from subjects enrolled in the RECOVERY study (ClinicalTrials.gov Identified: NCT02863354) [36]. For each subject, there is one montage FA image (Fig. 1(a)) and a set of longitudinal raw FA images (Figs. 1(b) and (c)) that were

taken nominally 24 weeks apart. The montage FA serves as the common reference image and the raw FAs are target images. All images are acquired using Optos California and 200Tx cameras (Optos plc, Dunfermline, United Kingdom) [37] and are stored in the TIFF format. The montage image, which was created by the Optos software via an equatorial stereographic projection of select individual images [26], [27], has a resolution of 4000×4000 whereas the individual FA images are 3900×3072 . For quantitative evaluation, we manually select a set of corresponding points, which we will refer to as control points, for each image pair. The control points are chosen to be the landmarks on the images, such as bifurcation points and endpoints of the retinal vessels. In addition, the locations of control points are chosen to provide coverage of the entire overlapping region between the two images.

The FIRE dataset consists of 134 pairs of narrow-field color fundus images [28]. All images have the same resolution of 2912×2912 and the same of field-of-view of 45° . The images are split into three categories with different characteristics. Category \mathcal{S} contains 71 image pairs that carry high overlap and no anatomical changes; Category \mathcal{A} contains 14 image pairs with large overlap and significant anatomical changes; Category \mathcal{P} includes 49 pairs of images with small overlap. For each image pair, 10 control points are manually labeled as the ground truth for evaluation.

B. Evaluation Metrics and Benchmark

We use the Mean Registration Error (MRE) [28], the area under the success rate curve, and the chamfer distance for quantitative evaluation. For each image pair, MRE is calculated as the average residual Euclidean distance (in pixels) between the control points under the estimated transformation. Given an MRE threshold τ_e , we define the success rate as the percentage of test image pairs in which the MRE is below the threshold τ_e

$$\text{SuccessRate}(\tau_e) = \frac{1}{N} \sum_i^N \mathbb{1}[\text{MRE}(\mathbf{I}_r^i, \mathbf{I}_{t \rightarrow r}^i) < \tau_e], \quad (6)$$

where N is the total number of image pairs in the test dataset and $\mathbb{1}$ is the indicator function. The success rate curve is plotted as a function of the MRE threshold τ_e . We compute the area under the curve (AUC) using the trapezoidal method and report the normalized AUC value in the range of $[0, 1]$.

In addition to the area under the success rate curve, we also report the residual chamfer distance (RCD) between the registered binary vessel maps, which provides a good proxy for the registration error. Since the chamfer distance is asymmetric, we compute the two-way chamfer distance by swapping the reference and the target images and report the median two-way chamfer distance computed over all test image pairs.

We consider the following methods as baselines for comparison that are designed for retinal image registration: REMPR [20], Harris-PIIFD [38], GFEMR [11], RIR-BS [39], SURF-PIIFD-RPM [40], and GAIN [41]. We use the code provided by the authors and keep the default parameter settings as suggested in the corresponding papers.

	Order of Polynomial	AUC \uparrow	RCD \downarrow
Global	2nd order	0.505	16.069
	3rd order	0.742	11.859
	4th order	0.719	14.462
Local	2nd order	0.840	5.127
	3rd order	0.854	5.100
	4th order	0.838	5.751

TABLE I: Evaluation of different orders of the polynomial transformation used in the global and the local registration steps on the FLoRI21 dataset.

C. Implementation, Ablation Study, and Analysis

We implement the proposed method using MATLABTM and PyTorch [42]. We use the pre-trained neural network models [32] and [43] to detect vessels from UWF FA images and narrow-field fundus images, respectively. Based on empirical evaluation, the feature matching threshold τ_1 is set to 0.5 and the inlier distance threshold τ_2 is set as 25. We run 800 RANSAC trials N_{trial} to find the best global transformation parameters. In the local registration stage, we convert the coarsely aligned images into 512×512 patches with a stride of 256 pixels. With this setting, the displacement field vector of each pixel location is estimated from 4 overlapping patches. Local registration cannot be reliably determined for patches with very few vessel pixels without using additional context. This situation is sometimes encountered for peripheral patches and addressed by combining patches having under 1% vessel pixels with the horizontally/vertically adjacent patch that has the larger number of vessel pixels.

To demonstrate the contribution of each stage in the proposed method and to verify the benefits of using vessel maps for the local registration, we conduct a series of ablation studies and algorithm analysis. We first provide an empirical justification for the orders of polynomial transformation used in the proposed hybrid registration framework. In Table I, we report the AUC and the ROC metrics for the second, third, and fourth order polynomial transformations in both global and local registration steps. For the global registration step, the third order polynomial transformation offers significant improvements over the second and the fourth order transformations. For the local registration, all the transformations achieve the similar performance with minor benefit from going to high-order polynomial transformation. Therefore, we use the third order and the second order polynomial transformations in the global and the local registration steps, respectively.

Next, we compare the proposed modified RANSAC matching algorithm with the traditional version on the FLoRI21 dataset. Note that the FLoRI21 dataset is more challenging than the FIRE dataset for RANSAC matching because the UWF FA images in the FLoRI21 dataset has significant geometric distortion in periphery. We implement two versions of the traditional RANSAC. The first one, denoted as RANSAC \mathcal{S}_c , estimates the global transformation parameters from the consensus set \mathcal{S}_c that contains the nearest neighbor for each SIFT [29] feature point. The second version, denoted as RANSAC \mathcal{S}_s , uses the sample set \mathcal{S}_s that includes only the

Method	#Inlier Points	AUC \uparrow	RCD \downarrow
RANSAC \mathcal{S}_c	220	0.170	19.798
RANSAC \mathcal{S}_s	407	0.703	13.662
Prop. RANSAC	1418	0.742	11.859

TABLE II: Quantitative comparison between the proposed RANSAC and the traditional RANSAC algorithms for estimating the global transformation parameters on the FLoRI21 dataset. “#Inlier Points” means the median number of inlier points determined by the RANSAC algorithms.

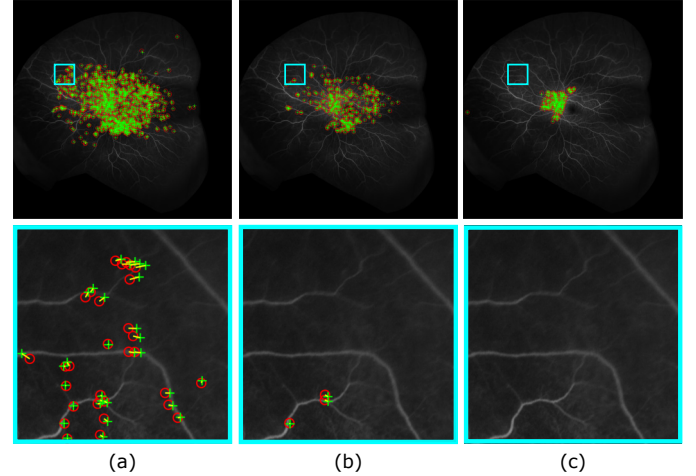


Fig. 4: Visual results of feature matching obtained from (a) the proposed RANSAC, (b) RANSAC \mathcal{S}_s , and (c) RANSAC \mathcal{S}_c . The red and green markers are the corresponding inlier SIFT features from the montage FA and the transformed raw FA, respectively.

distinctive SIFT feature matches. In addition to the AUC and RCD metrics, we also report the median number of matched inlier points estimated by RANSAC. Table II summarizes the results on the FLoRI21 dataset. The RANSAC \mathcal{S}_c finds the least number of inlier points because the tentative feature correspondences in \mathcal{S}_c contain significant mismatches. Therefore, RANSAC does not perform well on \mathcal{S}_c . While the median number of inlier points from RANSAC \mathcal{S}_c is 220, these matched features are not wide distributed over the image, as shown in Fig. 4(c). The RANSAC \mathcal{S}_s method can determine the transformation parameters that coarsely aligns image pairs. However, the number of matched inlier points is significantly smaller than the proposed RANSAC method (407 v.s. 1418, respectively). As shown in Fig. 4(b), RANSAC \mathcal{S}_s performs poorly in the peripheral regions with large geometric distortion. The proposed RANSAC matching method is able to robustly determine the inlier matched points by combining both the sets \mathcal{S}_c and \mathcal{S}_s .

In the third experiment, we demonstrate the effectiveness of the local registration step by comparing the proposed hybrid registration framework with global registration methods. We experiment with both hand-crafted SIFT [29] features and the learning-based SuperPoint [30] features. For SIFT feature extraction, we use the publicly available *VLFeat* library [44].

Global	Local	AUC \uparrow	RCD \downarrow
SIFT [29]	-	0.742	11.859
SuperPoint [30]	-	0.723	12.724
SIFT [29]	SIFT [29]	0.790	9.242
SIFT [29]	Diff Demons [45]	0.813	8.313
SIFT [29]	Optical Flow [46]	0.779	9.293
SIFT [29]	Intensity-MI [5]	0.793	9.432
SIFT [29]	Chamfer (Unsup. VD [47])	0.796	10.168
SIFT [29]	Chamfer (Sup. VD [32], Prop.)	0.840	5.127

TABLE III: Quantitative evaluation of the global and the local registration steps in the proposed method on the FLoRI21 dataset. Diff Demons: diffeomorphic demons; Unsup. VD: unsupervised vessel detection; Sup. VD: supervised vessel detection (proposed method). The methods in the first two rows only do not perform the local registration.

For SuperPoint [30] features, we adopt the pre-trained network provided with the publication¹. Table III reports the quantitative results on both the FLoRI21 and the FIRE datasets. Using the global registration as the initial step, the local registration step in the proposed hybrid framework, improves the registration accuracy significantly. On the FLoRI21 dataset, the proposed framework improves the AUC metric from 0.802 to 0.866 and reduces the RCD error from 8.859 to 5.127 pixels. A similar gain is also obtained on the FIRE dataset: the proposed framework achieves an AUC of 0.894 and the RCD of 1.730 pixels. We also note that the coarse alignment from the SuperPoint features, although slightly worse than that from the SIFT features, still provides good initialization for the local registration step.

In the fourth experiment, we justify the choices of the binary vessel map for the local registration by comparing the proposed method against alternative keypoint-based and intensity-based registration techniques for the second stage. To ensure the same experiment setting, we use the patch size of 512×512 pixels for all methods and perform the SIFT-based global registration to coarsely align the input image pairs. For the keypoint-based method, we select the matched SIFT features that lie inside each patch to estimate the local transformation using (2). For methods that take the intensity image as input, we consider the non-parametric deformable registration (diffeomorphic demons [45]), the optic flow using the TV-L1 solver algorithm [46], and the intensity-based registration that maximizes the mutual information (MI) between the image pair [5]. Quantitative results on the FLoRI21 dataset are listed in the second block of Table III. Compared with the global registration (AUC of 0.742 and RCD of 11.859), both feature-based and intensity-based local registration methods improves the registration accuracy (AUC of 0.813 and RCD of 8.313 for the best performing diffeomorphic demons method), which further reinforce the advantages of the two-stage hybrid registration framework. The proposed method, which uses binary vessel map in the local registration step, outperforms all alternative methods.

Lastly, we highlight that the local registration step in the proposed method benefits from the recent advances deep

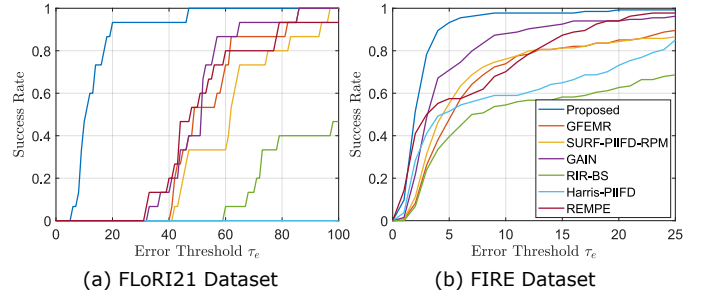


Fig. 5: Success rate curves for different retinal image registration methods on (a) the FLoRI21 dataset and (b) the FIRE dataset.

learning based vessel detection. Prior work in [4] adopts an unsupervised vessel detection method, MSMA [47], for the chamfer alignment. For comparison, we replace the pre-trained network model [32] in our implementation with MSMA and test the registration accuracy. Both methods are initialized using the feature-based global registration. Quantitative comparison on the FLoRI21 dataset, listed in the penultimate row in Table III, show that the registration using unsupervised vessel detection performs worse than that using deep neural network model. The unsupervised vessel detection method fails to detect many vessel branches from the FA image, especially in the low contrast region in the periphery. In contrast, the deep neural network can robustly detect both major and minor vessels, regardless of image contrast. An accurate vessel segmentation map is crucial in the local registration of the proposed method because the parametric chamfer alignment relies on binary vessel maps for anchoring. It is also worth noting that prior work in [4] directly applied the chamfer alignment on the input image, without the global registration step, which tends to be non-robust because the chamfer alignment is only locally convergent.

D. Results and Benchmarking

Figure 5 shows the success rate curves for different retinal image registration methods on both the FLoRI21 dataset and the FIRE dataset. Table IV summarizes the corresponding area under the curve (AUC) and the RCD metrics. On both datasets, the proposed hybrid registration method shows significant improvement over existing methods. The proposed method achieves an AUC of 0.866 and a RCD of 5.127 pixels on the FLoRI21 dataset and 0.894 AUC and 1.730 pixels RCD on the FIRE dataset. We observe that the performance of existing methods on the FLoRI21 datasets is much worse than those on the FIRE dataset. The primary reason for the poor performance is that these methods do not account for the significant distortion in the peripheral regions. For instance, the REMPR method [20] assumes that retinal camera can be modeled as a pinhole camera model without lens distortion. This assumption, however, is not valid for UWF retinal images.

Comparing the results presented in Table III, we notice that, on the FLoRI21 dataset, the proposed approach still performs better than existing methods even using alternative

¹<https://github.com/magicLeap/SuperPointPretrainedNetwork>

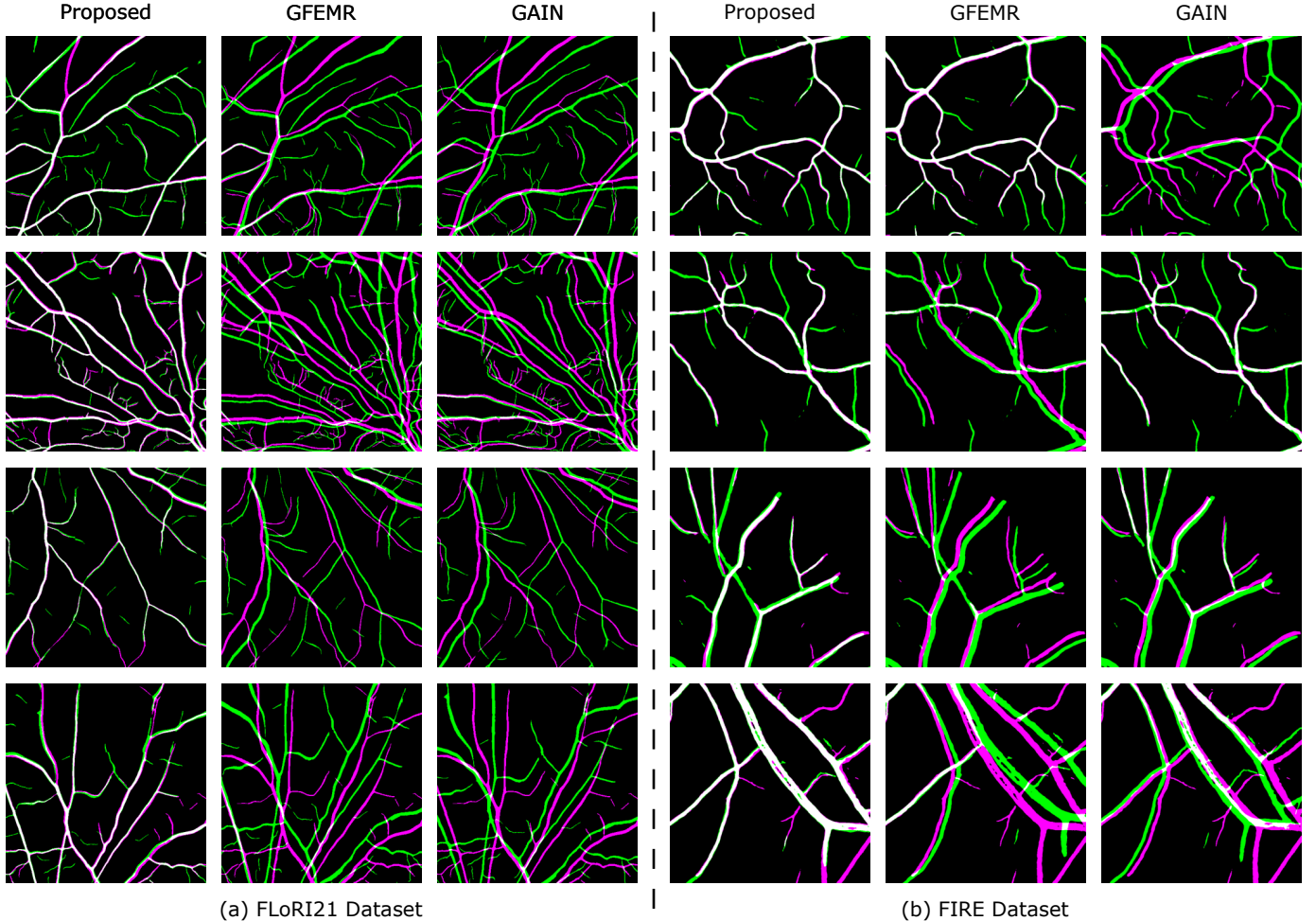


Fig. 6: Sample registration results for the proposed method and alternative methods on (a) the FLoRI21 dataset and (b) the FIRE dataset. The images are the composition of the reference vessel map (magenta) and the registered target vessel map (green). For a better visualization, we show the registered vessels on local patch level and provide the entire image in the Supplementary Material.

local registration techniques. The results further demonstrate the benefits of the proposed hybrid registration framework.

Figure 6 shows sample registered binary vessel maps obtained from different methods on the FLoRI21 dataset and the FIRE dataset. For a better visualization, we only include local vessel patches in Fig. 6 and provide the entire registered vessel images in Section S.IV in the Supplementary Material. The proposed hybrid registration method can accurately register the retinal images. For existing methods, we can see clear misalignment between the reference vessel map (magenta) and the registered target vessel map (green).

IV. CONCLUSION AND DISCUSSION

The novel two-stage hybrid registration framework proposed in this paper provides a useful approach for registering UWF retinal images to standardized stereographic projections where extreme geometric distortions are commonly encountered. Experimental evaluations on two datasets, including a new FLoRI-21 UWF dataset [35], show that the proposed approach outperforms prior retinal registration methods by a significant

Method	FLoRI21 Dataset		FIRE Dataset	
	AUC \uparrow	RCD \downarrow	AUC \uparrow	RCD \downarrow
Harris-PIIFD [38]	-	-	0.589	3.774
RIR-BS [39]	0.131	17.626	0.497	4.460
GAIN [41]	0.491	13.557	0.791	2.283
SURF-PIIFD-RPM [40]	0.372	16.029	0.680	2.420
GFEMR [11]	0.454	14.507	0.666	2.755
REMPR [20]	0.427	14.240	0.773	3.342
Proposed	0.840	5.127	0.894	1.730

TABLE IV: Comparative evaluations for different retinal vessel registration methods on the FLoRI21 and the FIRE datasets. The best result is shown in bold.

margin. The FLoRI21 dataset is made publicly available [35]² to facilitate further research on retinal image registration.

The second stage in the proposed framework effectively leverages the advances made in vessel segmentation and

²A sample pair of low resolution images is currently provided and the entire dataset will be released with the publication of the paper.

focuses on the anatomical features of interest, which is well matched with clinical scenarios where the longitudinal changes in vessel structures are of interest. Although beyond the scope of the present work, the idea could also be generalized to other image modalities, for instances, leveraging anatomical segmentation to register longitudinal brain MRI images [48].

ACKNOWLEDGMENT

The authors would like to thank the Center for Integrated Research Computing at the University of Rochester for providing computational resources and technical support.

REFERENCES

- [1] Early Treatment Diabetic Retinopathy Study Research Group, "Early photocoagulation for diabetic retinopathy: ETDRS report number 9," *Ophthalmology*, vol. 98, no. 5, Suppl, pp. 766 – 785, 1991.
- [2] H. Narasimha-Iyer *et al.*, "Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1436–1445, 2007.
- [3] K. M. Adal *et al.*, "An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1382–1390, 2018.
- [4] L. Ding *et al.*, "Quantification of longitudinal changes in retinal vasculature from wide-field fluorescein angiography via a novel registration and change detection approach," in *Proc. IEEE Intl. Conf. Acoustics Speech and Sig. Proc.*, Apr. 2018, pp. 1070–1074.
- [5] D. Mattes *et al.*, "Nonrigid multimodality image registration," in *SPIE Medical Imaging*, vol. 4322, 2001, pp. 1609–1620.
- [6] P. A. Legg *et al.*, "Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation," *Comput. Med. Imaging and Graphics*, vol. 37, no. 7-8, pp. 597–606, 2013.
- [7] K. M. Adal *et al.*, "A hierarchical coarse-to-fine approach for fundus image registration," in *Biomed. Image Registration*, 2014, pp. 93–102.
- [8] G. M. Lee *et al.*, "Unsupervised learning model for registration of multi-phase ultra-widefield fluorescein angiography," in *Intl. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2020, pp. 201–210.
- [9] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Proc.*, vol. 5, no. 8, pp. 1266–1271, 1996.
- [10] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. Assoc. Comp. Mach.*, vol. 24, no. 6, pp. 381–395, 1981.
- [11] J. Wang *et al.*, "Gaussian field estimator with manifold regularization for retinal image registration," *Signal Processing*, vol. 157, pp. 225–235, 2019.
- [12] S. K. Saha *et al.*, "A two-step approach for longitudinal registration of retinal images," *J. Med. Syst.*, vol. 40, no. 12, pp. 1–13, 2016.
- [13] J. A. Lee *et al.*, "A deep step pattern representation for multimodal retinal image registration," in *IEEE Intl. Conf. Comp. Vision.*, 2019, pp. 5077–5086.
- [14] P. Truong *et al.*, "GLAMPoints: Greedily learned accurate match points," in *IEEE Intl. Conf. Comp. Vision.*, 2019, pp. 10731–10740.
- [15] B. Fang and Y. Y. Tang, "Elastic registration for retinal images based on reconstructed vascular trees," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1183–1187, 2006.
- [16] A. M. Broehan *et al.*, "Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10, pp. 2816–2824, 2011.
- [17] E. Serradell *et al.*, "Non-rigid graph registration using active testing search," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 37, no. 3, pp. 625–638, 2015.
- [18] D. Motta, W. Casaca, and A. Paiva, "Vessel optimal transport for automated alignment of retinal fundus images," *IEEE Trans. Image Proc.*, vol. 28, no. 12, pp. 6154–6168, 2019.
- [19] Y. Lin and G. Medioni, "Retinal image registration from 2D to 3D," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2008, pp. 1–8.
- [20] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "REMPE: Registration of retinal images through eye modelling and pose estimation," *J. Biomed. and Health Informatics*, vol. 24, no. 12, pp. 3362–3373, 2020.
- [21] L. Ding *et al.*, "Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning," *IEEE Trans. Med. Imaging*, vol. 40, no. 10, pp. 2748–2758, Oct. 2021.
- [22] C. C. Wykoff *et al.*, "Progressive retinal nonperfusion in ischemic central retinal vein occlusion," *Retina*, vol. 35, no. 1, pp. 43–47, 2015.
- [23] M. H. Bawany *et al.*, "Automated vessel density detection in fluorescein angiography images correlates with vision in proliferative diabetic retinopathy," *PLOS ONE*, vol. 15, no. 9, pp. 1–14, Sept. 2020.
- [24] G. Yu *et al.*, "Quantification of Retinal Nonperfusion and Neovascularization With Ultrawidefield Fluorescein Angiography in Patients With Diabetes and Associated Characteristics of Advanced Disease," *JAMA Ophthalmology*, vol. 138, no. 6, pp. 680–688, 06 2020.
- [25] M. T. Witmer and S. Kiss, "Wide-field imaging of the retina," *Survey of Ophthalmology*, vol. 58, no. 2, pp. 143 – 154, 2013.
- [26] D. E. Croft *et al.*, "Precise montaging and metric quantification of retinal surface area from ultra-widefield fundus photography and fluorescein angiography," *Ophthalmic Surgery, Lasers and Imaging Retina*, vol. 45, no. 4, pp. 312–317, 2014.
- [27] *Digital Imaging and Communications in Medicine (DICOM) Supplement 173: Wide Field Ophthalmic Photography Image Storage SOP Classes*, DICOM Standards Committee, Rosslyn, VA, 2015, accessed Aug. 2021. [Online]. Available: <http://www.dclunie.com/dicom-status/status.html>
- [28] C. Hernandez-Matas *et al.*, "FIRE: fundus image registration dataset," *Modeling and Artif. Intel. in Ophthalmology*, vol. 1, no. 4, pp. 16–28, 2017.
- [29] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [30] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog. Wksp.*, 2018.
- [31] H. G. Barrow *et al.*, "Parametric correspondence and chamfer matching: Two new techniques for image matching," in *Proc. Int. Joint Conf. Artificial Intell.*, 1977, pp. 659–663.
- [32] L. Ding *et al.*, "A novel deep learning pipeline for retinal vessel detection in fluorescein angiography," *IEEE Trans. Image Proc.*, vol. 29, no. 1, pp. 6561–6573, 2020.
- [33] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B (methodol.)*, pp. 1–38, 1977.
- [34] G. Borgefors, "Distance transformations in digital images," *Comp. Vis., Graphics and Image Proc.*, vol. 34, no. 3, pp. 344–371, 1986.
- [35] L. Ding *et al.*, "FLoRI21: Fluorescein angiography longitudinal retinal image registration dataset," IEEE Dataport, 2021. [Online]. Available: <https://dx.doi.org/10.21227/ydp8-zf19>
- [36] C. C. Wykoff *et al.*, "Intravitreal aflibercept for retinal nonperfusion in proliferative diabetic retinopathy: Outcomes from the randomized RECOVERY trial," *Ophthalmology Retina*, vol. 3, no. 12, pp. 1076–1086, 2019.
- [37] *Optos California Tech Sheet*, Optos, 2015. [Online]. Available: <https://www.optos.com/globalassets/www.optos.com/products/c/california/california-brochure.pdf>
- [38] J. Chen *et al.*, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, 2010.
- [39] L. Chen *et al.*, "Retinal image registration using bifurcation structures," in *IEEE Intl. Conf. Image Proc.*, 2011, pp. 2169–2172.
- [40] G. Wang *et al.*, "Robust point matching method for multimodal retinal image registration," *Biomedical Signal Processing and Control*, vol. 19, pp. 68–76, 2015.
- [41] D. Motta, W. Casaca, and A. Paiva, "Fundus image transformation revisited: Towards determining more accurate registrations," in *Intl. Symp. Comput.-Based Medical Sys.*, 2018, pp. 227–232.
- [42] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Adv. in Neural Info. Proc. Sys.*, 2019, pp. 8024–8035.
- [43] Z. Zhou *et al.*, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Med. Image Analysis*, 2018, pp. 3–11.
- [44] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," <http://www.vlfeat.org/>, 2008.
- [45] T. Vercauteren *et al.*, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
- [46] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Joint Patt. Recog. Symp.*, 2007, pp. 214–223.

- [47] L. Ding *et al.*, “Multi-scale morphological analysis for retinal vessel detection in wide-field fluorescein angiography,” in *Proc. IEEE Western NY Image and Signal Proc. Wksp. (WNYISPW)*, Rochester, NY, Nov. 2017, pp. 1–5.
- [48] G. Balakrishnan *et al.*, “Voxelmorph: A learning framework for deformable medical image registration,” *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

Supplementary Material for “Combining Feature Correspondence with Parametric Chamfer Alignment: Hybrid Two-Stage Registration for Ultra-Widefield Retinal Images”

Li Ding, Tony D. Kang, Ajay E. Kuriyan, Rajeev S. Ramchandran, Charles C. Wykoff, and Gaurav Sharma

S.I. OVERVIEW

This document provides Supplementary Material for the paper [1]. In Section S.II, we present the least squares estimation for the global transformation parameters β for different order polynomial transformation. In Section S.III, we provide details of the EM-based chamfer alignment algorithm used in the second stage of our proposed hybrid registration algorithm by deriving the expressions for posterior probabilities and the parameter updates. Finally, in Section S.IV, we show larger images corresponding to entire registered vessel maps for the proposed and alternative techniques; smaller regions from these are included in the main paper.

S.II. LEAST SQUARES PARAMETER ESTIMATES FOR THE GLOBAL REGISTRATION TRANSFORM

The polynomial transformation parameters β used for the global geometric transform in Section II-A of the main manuscript are obtained via a least-squares procedure that we outline in this section. The geometric transformation from the target to the reference spatial coordinates is defined in terms of a polynomial in the variables x and y , as

$$\psi_M([x, y]^T, \beta) = \sum_{m=0}^M \sum_{i=0}^m \beta_{m,i} x^{m-i} y^i \stackrel{\text{def}}{=} \beta_M^T \mathbf{a}([x, y]^T), \quad (\text{S.1})$$

$$\beta_M = [\beta_{0,0}, \beta_{1,0}, \beta_{1,1}, \beta_{2,0}, \beta_{2,1}, \dots, \beta_{M,M}]^T, \quad (\text{S.2})$$

$$\mathbf{a}([x, y]) = [1, x, y, x^2, xy, \dots, y^M]^T, \quad (\text{S.3})$$

where M is the order of the polynomial and $\beta_{m,i}, m = 0 \dots M, i = 0 \dots m$ denote the $(M+1)(M+2)/2$ coefficients of the polynomial. In terms of the polynomial in (S.1), we can then express the M -th order polynomial geometric transformation as

$$\bar{\mathcal{T}}_{\beta_M}(\mathbf{p}^{(t)}) = \beta_M^T \mathbf{a}(\mathbf{p}^{(t)}), \quad (\text{S.4})$$

where $\mathbf{p}^{(t)}$ and $\bar{\mathcal{T}}_{\beta_M}(\mathbf{p}^{(t)})$ are 2×1 coordinate vectors denoting corresponding spatial locations in the target and reference images, respectively, $\beta_M = [\beta_M^u, \beta_M^v]$ denotes the matrix whose columns β_M^u and β_M^v are the parameter vectors for M -th order polynomial transformations corresponding to the two spatial coordinates in the reference image space, denoted by u and v , respectively.

Given N pairs of matched feature points $\mathbf{p}_n^{(r)} = (u_n, v_n)^T$ and $\mathbf{p}_n^{(t)} = (x_n, y_n)^T$ in the reference and target images, respectively, we would like to determine the optimal transformation parameters β_M^* such that the transformed target coordinates closely approximate their corresponding reference points, i.e., $(\beta_M^*)^T \mathbf{a}_n \approx \mathbf{p}_n^{(r)}$ for $n = 1, 2, \dots, N$. Least squares estimates are readily obtained as

$$\beta_M^* = \arg \min_{\beta_M} \|\mathbf{A}\beta_M - \mathbf{b}\|^2 = \mathbf{A}^\dagger \mathbf{b}, \quad (\text{S.5})$$

The work was supported in part by a University of Rochester Research Award, by a distinguished researcher award from the New York state funded Rochester Center of Excellence in Data Science (CoE #3B C160189) at the University of Rochester, by an unrestricted grant to the Department of Ophthalmology from Research to Prevent Blindness, and grant P30EY001319-35 from the National Institutes of Health.

L. Ding and G. Sharma are with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627, USA (e-mail: {l.ding, gaurav.sharma}@rochester.edu).

T. D. Kang and R. S. Ramchandran are with the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: {tony_kang, rajeev_ramchandran}@urmc.rochester.edu).

A. E. Kuriyan is with the Mid Atlantic Retina, Retina Service of Wills Eye Hospital, Thomas Jefferson University, Philadelphia, PA 19107 & the University of Rochester Medical Center, University of Rochester, Rochester, NY 14642, USA (e-mail: ajay.kuriyan@gmail.com).

C. C. Wykoff is with Retina Consultants of Houston and Blanton Eye Institute, Houston Methodist Hospital & Weill Cornell Medical College, Houston, TX 77030, USA (e-mail: ccwmd@houstonretina.com).

where

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]^\top, \quad (\text{S.6})$$

$$\mathbf{b} = [\mathbf{p}_1^{(r)}, \mathbf{p}_2^{(r)}, \dots, \mathbf{p}_N^{(r)}]^\top, \quad (\text{S.7})$$

and \mathbf{A}^\dagger denotes the pseudo-inverse of \mathbf{A} given by $\mathbf{A}^\dagger = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top$ for the typical situation where $(\mathbf{A}^\top \mathbf{A})$ is non-singular.

S.III. DETAILED DERIVATION OF THE PARAMETER ESTIMATION WITH THE EM ALGORITHM

We describe the computation of the registration parameters for the k^{th} pair of coarsely aligned corresponding reference and target image patches and therefore assume that k is fixed throughout the following description. For the j^{th} vessel pixel $\mathbf{p}_j^{(t,k)}$ in the k^{th} target image patch, the squared chamfer distance $d_j(\beta_k)$ to the corresponding reference image under the transformation $\tilde{\mathcal{T}}_{\beta_k}(\cdot)$ is defined as the minimum squared Euclidean distance between the transformed point $\tilde{\mathcal{T}}_{\beta_k}(\mathbf{p}_j^{(t,k)})$ and the set of vessel pixel locations $\mathcal{Q}_{r,k}$ for the corresponding reference image, i.e.,

$$d_j(\beta_k) = \min_i \left\| \mathbf{p}_i^{(r,k)} - \tilde{\mathcal{T}}_{\beta_k}(\mathbf{p}_j^{(t,k)}) \right\|^2. \quad (\text{S.8})$$

The E-step calculates the expectation of the complete-data log-likelihood

$$\begin{aligned} Q &= \mathbb{E} \left[\sum_{j=1}^{M_{t,k}} \log p(d_j, z_j^k | \pi_k, \lambda_k, \beta_k) \right] \\ &= \sum_{j=1}^{M_{t,k}} \sum_{z_j^k \in \{0,1\}} p(z_j^k | \pi_k, \lambda_k, \beta_k, d_j) \log p(d_j, z_j^k | \pi_k, \lambda_k, \beta_k) \\ &= \sum_{j=1}^{M_{t,k}} p_j^k [-\lambda_k d_j + \log(\pi_k) + \log(\lambda_k)] + (1 - p_j^k) [\log(1 - \pi_k) - \log(P)], \end{aligned} \quad (\text{S.9})$$

where p_j^k is the posterior probability given the current parameter estimates, which is obtained using Bayes' rule as

$$\begin{aligned} p_j^k &= p(z_j^k = 1 | \pi_k, \lambda_k, \beta_k, d_j) \\ &= \frac{\pi_k \lambda_k e^{-\lambda_k d_j}}{\pi_k \lambda_k e^{-\lambda_k d_j} + \frac{1 - \pi_k}{P}}. \end{aligned} \quad (\text{S.10})$$

In the M-step, we determine updated values for the parameters π_k , λ_k , and β_k by maximizing the expectation Q , which yields

$$\pi_k = \frac{\sum_{j=1}^{M_{t,k}} p_j^k}{M_{t,k}}, \quad \lambda_k = \frac{\sum_{j=1}^{M_{t,k}} p_j^k d_j}{\sum_{j=1}^{M_{t,k}} p_j^k}, \quad (\text{S.11})$$

and

$$\beta_k = \arg \min_{\beta} \frac{1}{M_{t,k}} \sum_{j=1}^{M_{t,k}} p_j^k d_j(\beta). \quad (\text{S.12})$$

We use the Levenberg-Marquardt (LM) [2] algorithm to solve the optimization problem in (S.12). Starting with an initial estimate of $\beta_k^{(0)}$ (chosen as the identity transformation in our local registration setting), the LM algorithm obtains the optimum on the right-hand-side of (S.12) as the limit of the sequence $\beta_k^{(l+1)} = \beta_k^{(l)} + \Delta$, where $\Delta = [\Delta^u, \Delta^v]$, with Δ^u and Δ^v representing the parameter increments for β_k^u and β_k^v , respectively. The parameter increment Δ^u is obtained by solving the linear equation

$$\left(\sum_{j=1}^{M_{t,k}} \mathbf{J}_j^\top \mathbf{J}_j + \sigma \mathbf{E} \right) \Delta^u = 2 \sum_{j=1}^{M_{t,k}} p_j^k r_{j,u}^k \mathbf{J}_j^\top, \quad (\text{S.13})$$

where $r_{j,u}^k$ is the u -component of the residual distance vector

$$\mathbf{r}_j^k = \min_i \left(\mathbf{p}_i^{(r,k)} - \tilde{\mathcal{T}}_{\beta_k}(\mathbf{p}_j^{(t,k)}) \right), \quad (\text{S.14})$$

which can be efficiently obtained from the distance transform [3] and \mathbf{J}_j is the Jacobian of the u -component of the transformation $\tilde{\mathcal{T}}_{\beta_k}(\mathbf{p}_j^{(t,k)})$ evaluated at the j^{th} target point $\mathbf{p}_j^{(t,k)}$, which is obtained as

$$\mathbf{J}_j = \frac{\partial (\mathbf{a}_j^\top \beta_k^u)}{\partial \beta_k^u} = \mathbf{a}_j^\top, \quad (\text{S.15})$$

\mathbf{E} is the identity matrix, and σ is an iteration dependent damping parameter. If the parameter increment obtained by solving (S.13) leads to an increase in error, then σ is multiplied by a factor of 10. Otherwise, σ is divided by the same factor. The increment Δ^v is similarly obtained.

S.IV. VISUAL RESULTS OF THE ENTIRE REGISTERED VESSEL MAP

We provide larger images corresponding to the entire registered vessel maps from which smaller regions were presented in the results of Fig. 6 in the main manuscript. Figures S.1 and S.2 show the images from the FLoRI21 [4] and the FIRE [5] datasets, respectively. For consistency, Figs. S.1 and S.2 use the same layout as Fig. 6 in the main manuscript, readers can view the PDF document under high zoom to see individual images or selected regions within the images.

REFERENCES

- [1] L. Ding *et al.*, “Combining feature correspondence with parametric chamfer alignment: Hybrid two-stage registration for ultra-widefield retinal images,” submitted for review.
- [2] J. Nocedal and S. Wright, *Numerical optimization*. New York, NY, USA: Springer, 2006.
- [3] G. Borgefors, “Distance transformations in digital images,” *Comp. Vis., Graphics and Image Proc.*, vol. 34, no. 3, pp. 344–371, 1986.
- [4] L. Ding *et al.*, “FLoRI21: Fluorescein angiography longitudinal retinal image registration dataset,” IEEE Dataport, 2021. [Online]. Available: <https://dx.doi.org/10.21227/ydp8-zf19>
- [5] C. Hernandez-Matas *et al.*, “FIRE: fundus image registration dataset,” *Modeling and Artif. Intel. in Ophthalmology*, vol. 1, no. 4, pp. 16–28, 2017.

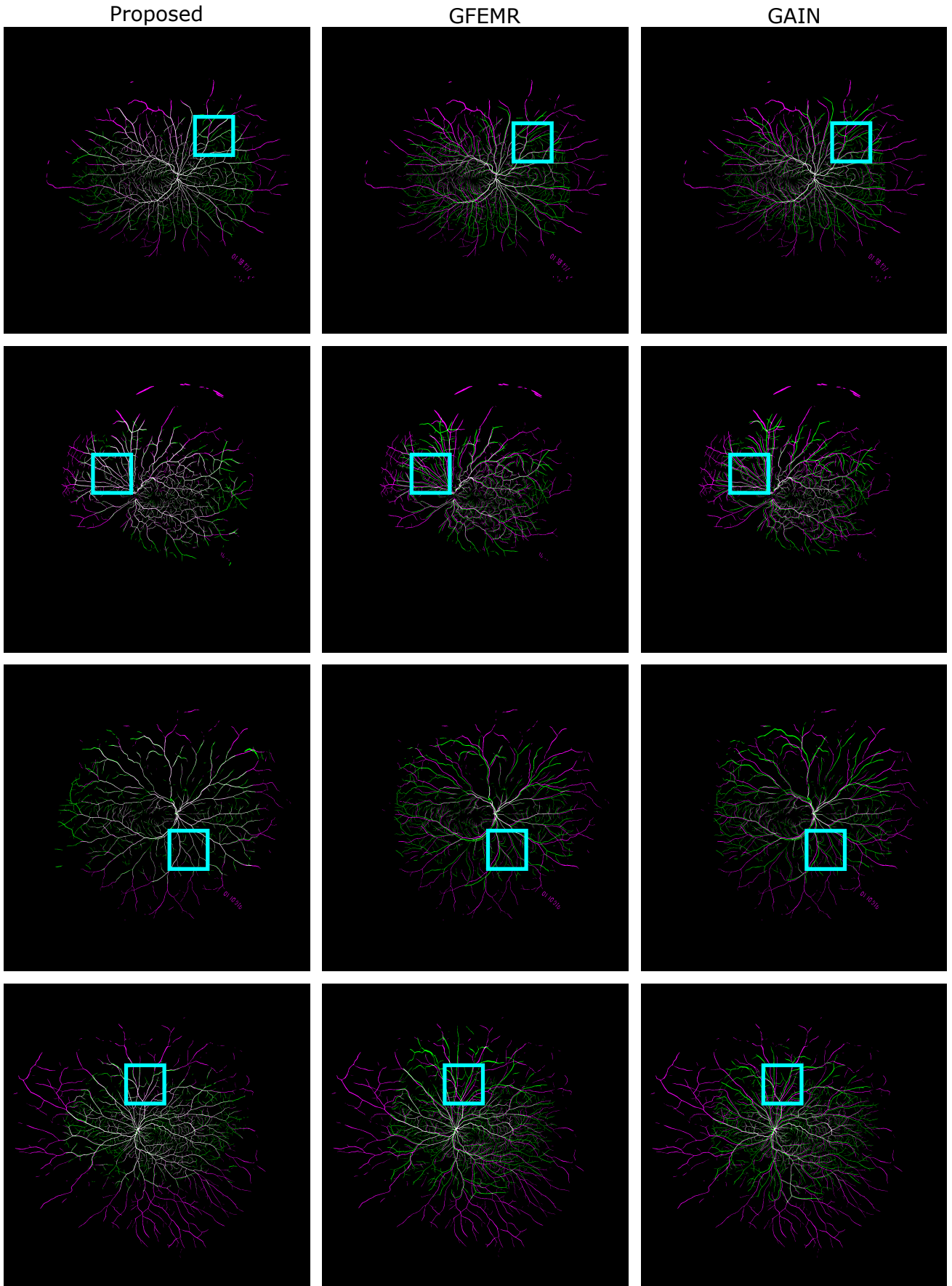


Fig. S.1: Sample registration results for the proposed method and alternative methods on the FLoRI21 dataset. The images are the composition of the reference vessel map (magenta) and the registered target vessel map (green). The regions highlighted by the cyan rectangles are shown in the Fig. 6(a) of the main manuscript.

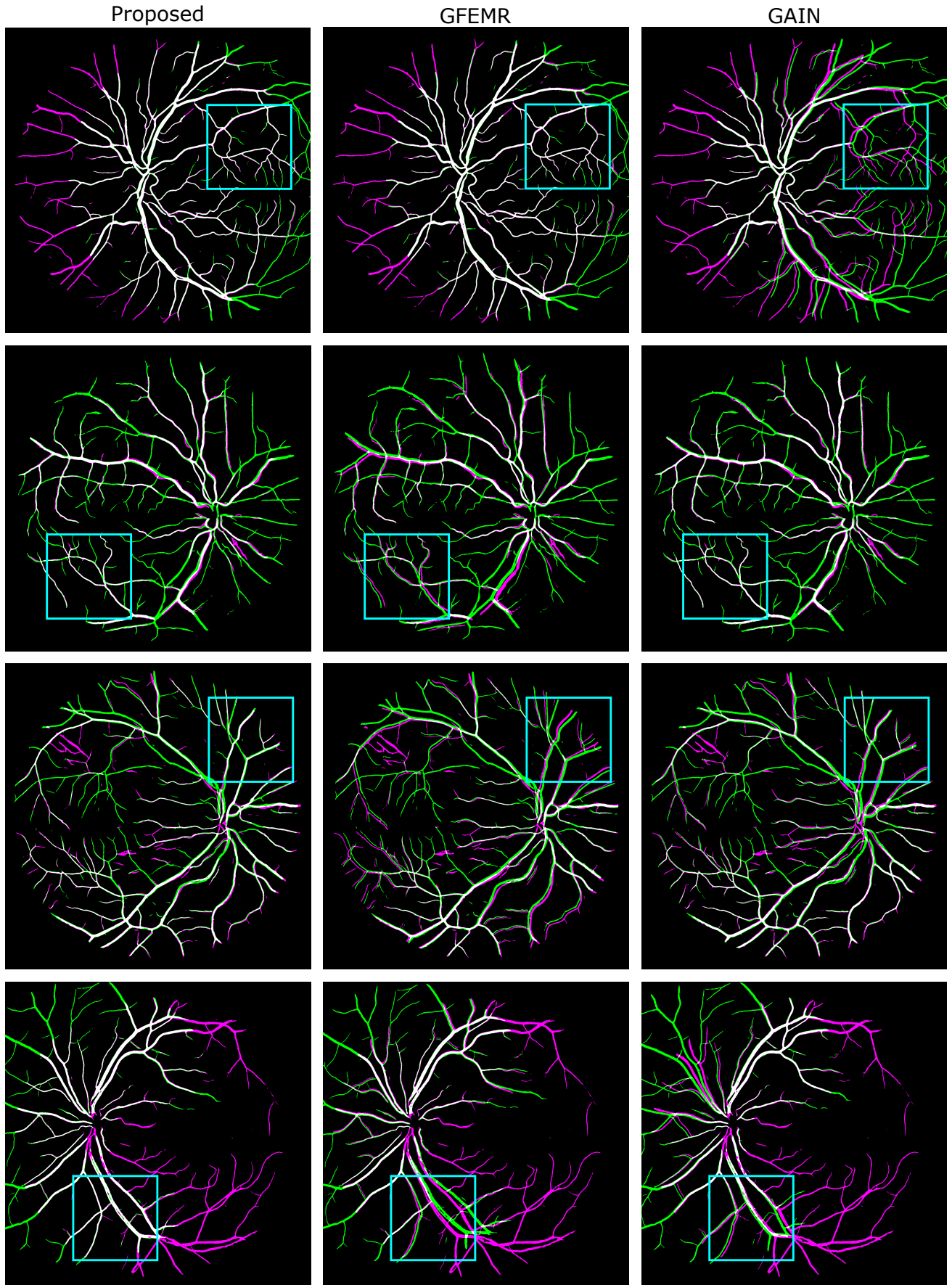


Fig. S.2: Sample registration results for the proposed method and alternative methods on the FIRE dataset. The images are the composition of the reference vessel map (magenta) and the registered target vessel map (green). The regions highlighted by the cyan rectangles are shown in the Fig. 6(b) of the main manuscript.