# REMPE: Registration of Retinal Images Through Eye Modelling and Pose Estimation

Carlos Hernandez-Matas<sup>®</sup>, Xenophon Zabulis<sup>®</sup>, and Antonis A. Argyros<sup>®</sup>

Abstract-Objective: In-vivo assessment of small vessels can promote accurate diagnosis and monitoring of diseases related to vasculopathy, such as hypertension and diabetes. The eye provides a unique, open, and accessible window for directly imaging small vessels in the retina with non-invasive techniques, such as fundoscopy. In this context, accurate registration of retinal images is of paramount importance in the comparison of vessel measurements from original and follow-up examinations, which is required for monitoring the disease and its treatment. At the same time, retinal registration exhibits a range of challenges due to the curved shape of the retina and the modification of imaged tissue across examinations. Thereby, the objective is to improve the state-of-the-art in the accuracy of retinal image registration. Method: In this work, a registration framework that simultaneously estimates eye pose and shape is proposed. Corresponding points in the retinal images are utilized to solve the registration as a 3D pose estimation. Results: The proposed framework is evaluated quantitatively and shown to outperform state-of-theart methods in retinal image registration for fundoscopy images. Conclusion: Retinal image registration methods based on eye modelling allow to perform more accurate registration than conventional methods. Significance: This is the first method to perform retinal image registration combined with eye modelling. The method improves the state-of-the-art in accuracy of retinal registration for fundoscopy images, quantitatively evaluated in benchmark datasets annotated with ground truth. The implementation of registration method has been made publicly available.

*Index Terms*—Image registration, medical image registration, medical imaging, retinal image registration, retinal imaging.

#### I. INTRODUCTION

MALL vessels exist in all organs of the human body, but the retina provides an easily accessible way to non-invasively

Manuscript received November 17, 2019; revised February 19, 2020; accepted March 23, 2020. Date of publication April 2, 2020; date of current version December 4, 2020. This work was supported in part by the Marie Curie grant from the EC in the framework of the REVAMMAD ITN (Initial Training Research Network), Project 316990 and in part by the FORTH-ICS internal RTD Programme "Ambient Intelligence and Smart Environments". (Corresponding author: Carlos Hernandez-Matas.)

Carlos Hernandez-Matas is with the Foundation for Research and Technology-Hellas, 70013 Heraklion, Greece, and also with the University of Crete, 70013 Heraklion, Greece (e-mail: carlos@hernandez.im).

Xenophon Zabulis is with the Foundation for Research and Technology-Hellas, 70013 Heraklion, Greece (e-mail: zabulis@ics. forth.gr).

Antonis A. Argyros is with the Foundation for Research and Technology-Hellas, 70013 Heraklion, Greece, and also with the University of Crete, 70013 Heraklion, Greece (e-mail: argyros@ics.forth.gr).

Digital Object Identifier 10.1109/JBHI.2020.2984483

estimate the microvascular status via fundoscopy [1]. This facilitates diagnosis and progression monitoring of diseases with strong vasculopathy, such as hypertension [2] and diabetes [3]. The analysis of retinal structures is also important for the diagnosis of illnesses that affect the eyesight [1].

Diagnosis and disease monitoring is facilitated by accurate retinal image registration. Image registration is applied upon a pair of images, the reference  $(F_0)$  and the test  $(F_t)$  one. The goal is the warping of  $F_t$  so that corresponding points in both images occur at the same 2D locations in the reference frame of  $F_0$ . Retinal Image Registration (RIR) [4] is challenging due to optical differences in various devices or modalities, anatomical changes (i.e., due to retinopathy), and acquisition artifacts. Viewpoint differences perplex image registration due to projective distortions in the images. Another complexity arises because of the curved shape of the retinal fundus which needs to be accounted for accurate registration.

Medical applications of RIR can be classified according to whether images are acquired during the same or in different sessions. Images acquired during the same examination typically lack significant anatomic changes. If the overlap in the pair of images is significant, they can be combined into super resolved images, i.e., images of higher resolution and definition [5]–[7] that enable more accurate measurements of the vessel structure such as Arteriolar-to-Venular Ratio [8]. Images with little overlapping surface can be utilized to create image mosaics that present larger retinal areas [9]–[11].

Longitudinal studies of the retina [12], [13] can be performed with registered images acquired at different points in time. These studies allow to monitor health status and disease progression. They provide an alternative method for the assessment of the effectiveness of a treatment and patient response. While differences due to retinopathy such as hemorrhages can be clearly identified without the assistance of additional tools, registration may prove useful for detecting minute changes such as differences in vasculature width, which is relevant for the study of hypertensive retinopathy.

This work proposes a novel RIR method. Compared to existing approaches, its novelty is that registration is achieved by considering and jointly estimating the relative pose of the cameras that acquired the images, as well as the parameters of an ellipsoidal model of the eye. A preliminary version of this method [14] presented fundamental image registration framework, utilizing Speeded Up Robust Features (SURF) features, Particle Swarm Optimization (PSO) and a fixed spherical model. The work in [15] expanded this method by (i) adding the Random

2168-2194 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information. Sample Consensus (RANSAC) initialization and (ii) utilizing Scale-Invariant Feature Transform (SIFT) instead of SURF features. The work in [16] focused uniquely on substituting the spherical model with ellipsoids to better approximate the eye shape. In [17] we studied the impact of utilizing diverse keypoints and their combinations. In this work we propose an alternative initialization step to our approach. Moreover, we present the world model geometry, as the basis of the proposed method. Additionally, we provide a consolidated and comprehensive presentation of all aspects of the proposed RIR framework. This presentation is followed by an in depth ablation analysis that is based on a publicly available and ground-truth annotated dataset. Access to the executable of the proposed method is also provided, in this submission.

#### II. RELATED WORK

Registration algorithms can be classified according to diverse criteria, such as the the type of information used, how such information is used, and which is the transformation model employed to align the images [4].

RIR methods based on the frequency domain [18], [19] only allow for 2D transformations. In the RIR case, spatial methods seem to be a better alternative as they allow for more complex transformations.

Within spatial methods, several utilize 2D transformations [20]–[29], or projective transformations [30], [31]. They provide accurate registration results for images with narrow field of view (FOV) or large overlap, but their performance decreases for images with larger FOV, where the curvature of the eye introduces distortion in the image.

Methods based on quadratic transformations are the ones that can achieve the most accurate registration, as they have the potential to approximate the 3D shape of the eye. Within methods that use quadratic transformations, the ones utilizing the whole image such as [32] are computationally expensive due to the large amount of information used. Methods utilizing localized features are typically faster to execute. SURF and Partial Intensity Invariant Feature Descriptor (PIIFD) based keypoints such as the ones used in [33]–[36] are shown [17] to provide less accurate registration results than other keypoint features such as SIFT [15]-[17] and vessel bifurcations [10], [17], [37]–[39]. RIR registration methods typically do not utilize simultaneously different types of features. A series of closely related methods based on Iterative Closest Point (ICP) [40]–[42] utilizing diverse features provide accurate registration results when a good initialization is performed. However, the fact that this step is based upon a single SIFT keypoint match proves to be a weakness in challenging image pairs [16].

Methods that utilize second order polynomials for the transformation, tend to obtain accurate registration for the overlapping areas in the images. However, the transformation performed in non-overlapping areas sometimes is not accurate. Such a drawback could be overcome by utilizing a model approximating the shape of an eye [15], [31]. State-of-the-art research efforts orbit towards extending already existing approaches such as RANSAC [43] or opening new approaches such as introducing



Fig. 1. Image acquisition and registration following the geometry presented in [16].

novel local descriptors for retinal images such as the D-Saddle feature [44] or leveraging the radial distortion of images [45]. However, similarly to what happens in several other computer vision domains, large efforts are targeted towards the utilization of Deep Neural Networks [46]–[48]. In particular, the method presented in [49] the Deep Step Patterns (DeepSPa) framework uses a Convolutional Neural Network (CNN) to register multimodal retinal images without the need for manually labelled datasets for training or manual intervention, focusing on the intensity change patterns rather than the intensity change values of the images, and showing great potential for this direction on the RIR field.

With respect to the aforementioned classifications, in this work, we propose a local, spatial method that performs retinal image registration by exploiting 3D eye models.

#### **III. REMPE REGISTRATION FRAMEWORK**

In fundoscopy, the camera typically has a fixed pose and a headrest is utilized to stabilize and align the head in front of the camera. Ideally, the eye is centered in front of the camera and its rotation avails views of diverse retinal areas to the camera. In this work, an equivalent geometry is considered: the reference camera that acquires the reference image  $(F_0)$  is fixed. Both eye rotation and pose of the camera that acquires the test image  $(F_t)$  are defined relative to the reference camera. The proposed approach, named Registration through Eye Modelling and Pose Estimation (REMPE), registers  $F_0$  and  $F_t$  by simultaneously estimating the relative pose of the eye when the retinal images were acquired, as well as the eye shape and rotation. Fig. 1 shows the aforementioned image acquisition and registration geometry. Point correspondences between the images are utilized to achieve this registration which is achieved by searching for the 3D geometry of camera postures and eye shape that best explain the 2D coordinates of matched points. An initial pose



Fig. 2. Workflow of REMPE.

estimate is calculated utilizing RANSAC and a spherical eye model. Subsequently, PSO is utilized to refine this pose, as well as to estimate the lengths of the semi-axes of the ellipsoidal eye model and their rotation. The workflow of the registration framework is shown in Fig. 2.

## A. Parameterization of Eye Shape and Camera Pose

The general shape of the eye can be represented via various geometrical models. In this work, an ellipsoidal model  $\{\mathbf{A}, \mathbf{Q}\}$  is utilized. Locating the coordinate origin at the center  $\mathbf{c}_s = [0, 0, 0]^T$  we obtain the ellipsoid equation

$$\mathbf{x}^T \mathbf{Q}^T \mathbf{A} \mathbf{Q} \mathbf{x} = 1, \tag{1}$$

where x is a point on the eye model surface  $\mathcal{E}$ . This model has 3 orthogonal semi-axes [a, b, c] composing A as shown in Eq. (2)

$$\mathbf{A} = \begin{bmatrix} a^{-2} & 0 & 0\\ 0 & b^{-2} & 0\\ 0 & 0 & c^{-2} \end{bmatrix},$$
 (2)

and rotations  $\mathbf{Q} = \mathbf{R}_a(r_a) \cdot \mathbf{R}_b(r_b) \cdot \mathbf{R}_c(r_c)$  of said semi-axes, leading to  $\mathcal{E}$ . A calibrated camera for  $F_0$  is located at  $\mathbf{c}_c = [0, 0, -\delta]^T$ .  $K_c$  and  $K_t$  are the intrinsic camera matrices for  $F_0$  and  $F_t$ .

A parameterization of eye shape and camera pose **S** in the world model consists of relative camera pose {**R**, **t**} and ellipsoid model {**A**, **Q**}, thus **S** = {**R**, **t**, **A**, **Q**}, totalling 12 parameters. Relative camera pose {**R**, **t**} corresponds to the relative rotation **R** = **R**<sub>x</sub>(r<sub>θ</sub>) · **R**<sub>y</sub>(r<sub>φ</sub>) · **R**<sub>z</sub>(r<sub>ω</sub>) and translation **t** = [t<sub>x</sub> t<sub>y</sub> t<sub>z</sub>]<sup>T</sup> between the cameras.

# B. World Model

The 3D "world" model of the scene and camera views is composed by two virtual acquisition devices and the eye model and can be represented by a few parameters: Lens to cornea distance (l) and Field of view (k) are known from the device specifications. Image radius, in pixels, (r), is availed by image size. An initial, coarse estimate of the eye model is chosen as per the Navarro eye model [50], that is a sphere ( $\rho = 12$  mm). Camera pixel spacing (p) is an intermediate variable that is eventually simplified in Eq. (5).



Fig. 3. World geometry.

Focal distance is calculated as follows:

$$f = p \ r \ \frac{l + \rho + \rho \ \cos\left(\frac{k}{2}\right)}{\rho \ \sin\left(\frac{k}{2}\right)}.$$
 (3)

A geometric representation of the framework is shown in Fig. 3. This geometry is central in this research, as it formulates the transfer of a fundus image from its 2D form to the 3D world, and guides its projection onto a 3D eye model and vice versa.

#### C. Intrinsic Camera Matrix

The intrinsic camera matrix K (Eq. (4)), together with the extrinsic parameters  $\{\mathbf{R}, \mathbf{t}\}$  of the camera allow to project a 3D point onto a camera view.

$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$
(4)

In this work, lens distortion is assumed to be negligible. In Eq. (4),  $\alpha_x$  and  $\alpha_y$  represent the focal length in pixels,  $\gamma$  is the skew coefficient between the x and y axes and  $u_0$  and  $v_0$  represent the center of the image, in pixels. The values of  $\alpha_x$  and  $\alpha_y$  are calculated from f in Eq. (3) as follows:

$$\alpha_x = \alpha_y = \frac{f}{p} = r \, \frac{l + \rho + \rho \, \cos\left(\frac{k}{2}\right)}{\rho \, \sin\left(\frac{k}{2}\right)}.$$
 (5)

#### D. Ray-Ellipsoid Intersection

Given **S**, the image locations are traced to 3D retinal locations upon the surface  $\mathcal{E}$  of the model {**A**, **Q**}. This is achieved by considering the visual ray from eye center  $\mathbf{c}_s$  through image keypoint **u**. Let point **x** be the intersection of this ray and  $\mathcal{E}$ . The 3D coordinates of **x** are given by:

$$\mathbf{x} = P^+ \mathbf{u} + \lambda \mathbf{c},\tag{6}$$

by solving the ray's equation (see [51, p. 162]):

$$P^{+} = P^{T} (PP^{T})^{-1} \tag{7}$$

for  $\lambda$ , where c is the camera center and P the projection matrix.

# E. Keypoint Correspondences

Keypoint detection and descriptor extraction methods have been widely utilized to identify common points in a pair of images. In this work, several methods used for RIR have been studied. A large set of RIR methods [15], [16], [31], [34], [41], [42] have been based on SIFT features [52]. SURF features [53] have also been employed [7], [14], [35], [36]. Harris-PIIFD [33] is a method introduced with the purpose of finding keypoints in cross-modal retinal image pairs. RIR methods employing PIIFD are found in [33]-[35]. Vessel bifurcations, or Y-features, are also commonly used in RIR [10], [17], [21], [24], [27], [37], [40], [54]. In this work, bifurcations are extracted conventionally as in [17].

Matching of features is performed with the conventional bilateral method proposed by Lowe [52]. Only symmetric matches are kept.

Diverse types of keypoints may predominantly occur in different areas of an image, conveying complementary information for registration. Feature combinations are expected to provide more correspondences in complementary locations on the retina and, in this way, reinforce the accuracy of results. Thus, in [17], we studied all possible combinations of the four types of keypoints in order to determine the one that performs best in terms of registration accuracy.

If the input images are not grayscale, the keypoints are calculated on the green channel of the images, as it provides higher contrast in retinal images than the blue and red channels [14].

#### F. Initialization

An initial candidate solution  $S_0$  is estimated by solving the Perspective-n-Point (PnP) [55] problem. Two different approaches were considered: RANSAC [55] and Posest [56].

RANSAC enables outlier detection and was introduced as a solution to the PnP problem. This method estimates the 3D pose of an object given a set of 2D-3D correspondences and the camera projection matrix P.

Posest [56] achieves model-based pose estimation of rigid objects. It calculates an initial pose estimation using a PnP solver embedded in a RANSAC framework and the redescending Mestimator sample consensus (MSAC) cost function. Then, a nonlinear refinement of this pose is performed iteratively with the Levenberg-Marquardt [57] algorithm.

Both methods rely on 2D-3D correspondences. As  $\{A, Q\}$ are unknown, we utilize the spherical model and the calibration of the cameras to retrieve the 3D location of the points based on Eq. (6). If no initialization method is used,  $\{\mathbf{R}, \mathbf{t}\}$  is initialized as  $\{I, 0\}$ .

#### G. Optimization

The estimation of S, is formulated as the solution of an optimization problem. Given a hypothesis  $S_h$  with  $\{R_h, t_h\}$ , the keypoints of the image are traced to 3D locations on the eye model surface as in Eq. (6). Points  $q_i$  are the 3D locations on  $\mathcal{E}$  of keypoints from  $F_0$ . Points  $\mathbf{p}_{i,h}$  are the 3D locations on  $\mathcal{E}$ of keypoints from  $F_t$  calculated from  $S_h$  (see Fig. 1). The 3D distances of corresponding keypoints on  $\mathcal{E}$  are:

$$d_{i,h} = |\mathbf{q}_i - \mathbf{p}_{i,h}|. \tag{8}$$

The minimization of distances  $d_{i,h}$  form the basis of the objective function  $o(\cdot)$  to be minimized. To increase robustness to spurious matches on  $o(\cdot)$ , a percentile of accumulated distances  $d_{i,h}$  is used:

$$o(\mathbf{S}_{\mathbf{h}}) = \sum_{\mathbf{j}} \mathbf{d}_{\mathbf{j},\mathbf{h}},\tag{9}$$

where *j* enumerates the smallest 80% values of  $d_{i,h}$ .

## H. Particle Swarm Optimization (PSO)

The 12D space of hypotheses around a initial solution S = $\{\mathbf{R}, \mathbf{t}, \mathbf{A}, \mathbf{Q}\}\$  is denoted as:

$$[\gamma_{\theta}, \gamma_{\phi}, \gamma_{\omega}, \mu_x, \mu_y, \mu_z, \mu_a, \mu_b, \mu_c, \gamma_a, \gamma_b, \gamma_c], \qquad (10)$$

for parameters

$$[r_{\theta}, r_{\phi}, r_{\omega}, t_x, t_y, t_z, a, b, c, r_a, r_b, r_c],$$
(11)

with  $r_{\theta} \in [-\gamma_{\theta}/2, \gamma_{\theta}/2]$  and correspondingly for the rest of dimensions. A grid-based search of this space to optimize Eq. (9) is computationally prohibitive. Besides, that would create a discretized version of the problem. Thus, the objective function of Eq. (9) is optimized via PSO [58], a stochastic, derivative-free optimization method successfully employed for pose estimation not only in this [14], [15] but also in other domains [59], [60].

PSO achieves optimization by iteratively improving a candidate solution given an objective function. The method utilizes a set of p particles, or candidate solutions, that evolve through g generations from an initial random position and velocity in a multidimensional search-space.

PSO can be configured with few parameters. The objective function to utilize is not required to have known derivatives, and it can be discontinuous or even cross-modal [61]. The required number of objective function evaluations  $(p \cdot q)$  is relatively low. These evaluations determine the computational budget of the optimization [61]. Allocating a small budget entails the risk of terminating the process prematurely, providing a poor pose estimate. On the other hand, an overly large budget may lead to additional computation without noticeable improvements in accuracy. Selecting a budget requires proper balancing of the trade-off between the accuracy and the speed of the method [61]. Given a particular budget, the final performance of the method is impacted by how the p particles are distributed across ggenerations.

#### I. Search Space Structure Variants

Aiming at accuracy and robustness, several method variants were considered to explore the solution search space. These variants consist of different PSO configurations with and without **RANSAC** initialization:

1) Coarse (C): Baseline method described in Section III-H. 2) Coarse-to-Fine (CF): PSO is executed twice, with the total budget split in half, and utilizing the same budget in each execution. For the second execution, the search space is a hypercube reduced in all 12 dimensions and centered at the solution provided by the first execution.

3) RANSAC (R): In this approach,  $\{R_0, t_0\}$  constitutes the solution, i.e., the RANSAC-based initialization described in Section III-F.

4) R-C: Coarse variant with RANSAC initialization.

*5) R-F:* Fine variant with RANSAC initialization. Similar to variant R-C, but with reduced search space.

6) R-CF: Coarse-to-Fine with RANSAC initialization.

Thus, variants C and CF are initialized at  $\{I, 0\}$ . Variants R-C and R-CF are identical to them, but initialized at  $\{R_0, t_0\}$ .

The search hypercube in Eq. (10) is set to

$$[4, 4, 4, 4, 4, 4, 8, 8, 8, 360, 360, 360]$$
(12)

for coarse PSO processes, and to

$$[2, 2, 2, 2, 2, 2, 4, 4, 4, 180, 180, 180]$$
(13)

for fine PSO processes. These values do not depend on the type of images and their characteristics, and have been selected empirically based on the analysis of typical scenarios for retinal image acquisition. For coarse search, the limits are set more broadly than a typical scene allows for, providing the whole range of possible solutions and allowing to find a candidate in the vicinity of the optimal solution. The parameters for the fine search limit more drastically the search space, allowing a refined search around the initial hypothesis. The results presented in Section IV-F lead to selecting the variant R-F as the proposed one.

#### J. Eye Model Variants

We consider four eye models that approximate either the full shape of the eye, or part of it. All of them present a smooth surface. In increasing order of complexity, these models are:

1) *Plane:* Baseline representation of the retinal surface by a plane. While this may be appropriate for a reduced FOV, the approximation deviates when images for large FOVs. [a, b, c] = [10000, 10000, 1] and  $[r_a, r_b, r_c] = [0, 0, 0]$ .

2) Sphere: The spherical model consists of an ellipsoid with equal axes that is invariant to rotations about them. Thus  $[a, b, c] = [\rho, \rho, \rho]$  and  $[r_a, r_b, r_c] = [0, 0, 0]$  and  $\rho = 12 \text{ mm}$  as per [50].

3) Fixed Orientation Ellipsoid: A general ellipsoid of fixed orientation so that its semi-axes a, b, c are parallel to axes xx', yy' and zz', respectively. Thus  $[a, b, c] = [\rho_a, \rho_b, \rho_c]$  and  $[r_a, r_b, r_c] = [0, 0, 0]$ .

4) Ellipsoid: A general ellipsoid that may be rotated relative to the camera coordinate system. Thus  $[a, b, c] = [\rho_a, \rho_b, \rho_c]$  and  $[r_a, r_b, r_c] = [\theta_a, \theta_b, \theta_c]$ .

The higher the complexity, the higher the potential of the model to correctly approximate the eye shape and orientation, thus leading to more accurate registration results. While most retinal diseases do not severely affect the general eye shape, in cases like acute age-related macula degeneration, where drusen is accumulated under the retinal pigment epithelium, these changes might be noticeable [1]. The proposed method currently does not account for this behaviour, but it can be extended in that direction by utilizing two eye models, one for each of the images and simultaneously calculating their eye shape parameters.

#### K. Multiple Process Execution

Both RANSAC and PSO are of stochastic nature, so the proposed method is non-deterministic. This can lead to suboptimal results, as there exists the risk of entrapment in local minima. Due to this, the process comprised by RANSAC and PSO is executed s times, denoted as swarms, and the parameters that give rise to the best score in the objective function minimization are selected as the solution. Execution of multiple swarms leads to an increase on the computational cost, but it makes suboptimal solutions less probable to occur. Alternatives such as increasing the PSO budget, are found in the experiments not to be as effective in avoiding these local minima entrapment.

#### L. Data Output and Image Formation

Data can be output in several ways. When the solution **S** is chosen, Eq. (6) is utilized to estimate the locations of the pixels from  $F_t$  on the surface  $\mathcal{E}$  of the eye model. This allows to output the 3D coordinates and the color for each of the points, generating a 3D model of the retinal eye shape.

These locations are projected to the reference camera utilizing the projection matrix P. This allows to output a list of points with the 2D floating point locations and the color data of the points. Such data are useful for generating Super Resolution (SR) images when combined with similar data from other registered images from the same eye.

Finally, 2D floating point and color information is utilized to create an image utilizing bilinear interpolation, which is then output as an image file. Additionally, **S** is stored, allowing to recreate the output data.

## **IV. EXPERIMENTS**

The goal of the experiments is twofold. The first goal is to investigate how to configure all the elements from the REMPE framework to form an accurate and robust RIR method. Such experiments include investigating the appropriate keypoints, eye model, search space structure variant and PSO budget distribution. The second goal is to perform a quantitative and comparative evaluation of the proposed approach to state of the art methods in RIR.

While no explicit comparison is performed with previous versions of the REMPE method [14]–[17], each of their configurations are included in the evaluation of the different elements presented in the experiments in this section.

#### A. Experimental Setup

Experiments were run in an Intel Core i7-4770 CPU @ 3.40 GHz with 16 GB of RAM memory and a NVIDIA GeForce GT 730 on Windows 7 Professional. The proposed RIR framework was implemented in C++ with openCV and CUDA tools.

The best trade-off between computational cost and accuracy is achieved when utilizing the following configuration: Information is retrieved from the image pairs using a combination of SIFT and vessel bifurcations. Solution **S** is calculated with s = 3independent swarms consisting on a RANSAC initialization using the spherical eye model, followed by a fine PSO search



Fig. 4. Workflow of the proposed registration method.

utilizing the ellipsoidal eye model. The fine PSO search is performed on a cube with the dimensions indicated in Eq. (12). For each PSO process, a budget p/g = 10 k/300, is used. This configuration is utilized throughout the experiments, unless otherwise noted. A diagram of the method is shown in Fig. 4.

#### B. Dataset

The Fundus Image Registration (FIRE) dataset [62] is used for the experiments. It consists of a collection of 134 image pairs split into 3 categories according to image pairs characteristics. Category S contains 71 image pairs with high overlap and no anatomical changes. Category P contains 49 image pairs with small overlap and no anatomical changes. Category A contains 14 image pairs with high overlap and large anatomical changes.

#### C. Registration Evaluation

Evaluation of the registration accuracy is performed as in [62]. More specifically, the registration error in pixels for each image pair is calculated as the average error on 10 correspondences, distributed over the part of the retina that is commonly visible in the images to be registered. These correspondences, provided by [62], were manually selected by an expert and mainly consist on vessel bifurcations and vessel crossings, further refined using correlation of the images patches around them to diminish any possible bias effect due to human annotation. Results are shown using a 2D plot in which the x axis indicates an error threshold. If the error in registration accuracy of an image pair is below this threshold, the registration is considered as successful. The y axis of the plot corresponds to the percentage of successfully registered image pairs for a given threshold. This way, the accuracy of a method is monitored over a wide range of target registration accuracy and not based on an arbitrarily selected threshold. This measure facilitates the fair comparison of RIR methods, allowing a registration method to be selected based on the accuracy requirements of the intended application. We use such plots to show registration accuracy for each individual category, as well as for the whole FIRE dataset. Additionally, the Area Under Curve (AUC) is provided.

#### D. Keypoint Selection

We compare the impact of the keypoints described in Section III-E on the accuracy of the proposed RIR framework. We computed the AUC for every feature combination when

TABLE I PERCENTAGE OF AUC IN EACH FIRE CATEGORY FOR KEY KEYPOINT COMBINATIONS. × INDICATES A KEYPOINTS TYPE USED FOR REGISTRATION. THE RESULTS OF THE EXPERIMENTS FOR ALL COMBINATIONS CAN BE FOUND IN [17]

SIFT	SURF	PIIFD	Bif	S	$\mathcal{P}$	$\mathcal{A}$	FIRE
×				0.945	0.443	0.577	0.721
	×			0.947	0.348	0.466	0.675
		×		0.846	0.134	0.429	0.538
			×	0.953	0.516	0.563	0.751
×			×	0.958	0.541	0.660	0.773
×		×	×	0.959	0.490	0.657	0.754

TABLE II AVERAGE REGISTRATION ERROR ON SYNTHETIC IMAGES FOR RANSAC AND POSEST

Method	Sphere Ideal	Sphere Noisy	Ellipsoid Ideal	Ellipsoid Noisy
RANSAC	78.784	80.137	84.634	85.065
Posest	78.800	90.079	84.647	91.289

registering the FIRE dataset [62]. The results (see Fig. 5 in [17]) show that the combination of SIFT features and Bifurcations provides the most accurate registration results. This is because keypoints that provide complementary information increase the accuracy of the approach. Bifurcations are interest points located on the vessels, which provide information about the retinal structures of the image. SIFT responds to more generic interest points that are not located necessarily along retinal vessels.

#### E. Initialization

Two experiments aim to compare the performance of RANSAC [55] and Posest [56] as initialization strategies. For both methods, the spherical eye model is used for registration, as 2D-3D correspondences are required, and the spherical model is the most complex model for which  $\{A, Q\}$  is known.

In the first experiment, registration has been performed on two sets of 10 synthetic retinal images, each. One set was generated assuming a spherical eye, and the other on ellipsoid eyes with arbitrary rotation. For each image pair, three sets of correspondences were created. The *GT* set contains 500 pairs of perfect correspondences to be used as ground truth to measure registration error. The *Ideal (I)* set contains another 500 pairs of perfect correspondences, different than those in *GT*, to be used as keypoints for registration. Finally, the *Noisy (N)* set involves the correspondences in *I*, contaminated with uniform noise in the range [-25, 25] pixels that has been added to feature coordinates. Error has been measured by calculating the average distance of the corresponding ground truth points in each image pair after registration, in pixels, and then the average error over the whole set of image pairs.

The obtained results are shown in Table II. It can be verified that the performance of RANSAC and Posest is comparable when ideal data are employed, regardless of the actual eye shape. However, RANSAC performs considerably better in the case of noisy observations. This is explained as follows. RANSAC is a purely randomized search algorithm. As such, it explores



Fig. 5. Registration success for RANSAC and posest with a spherical model. The x axis marks, in pixels, the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.



Fig. 6. Registration success for for the 6 search space variants. The x axis marks, in pixels, the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.

TABLE III PERCENTAGE OF AUC FROM FIG. 5 IN EACH FIRE CATEGORY FOR RANSAC AND POSEST WITH A SPHERICAL MODEL

Method	S	$\mathcal{P}$	$\mathcal{A}$	FIRE
RANSAC	0.933	0.209	0.549	0.624
posest	0.919	0.191	0.423	0.598

solutions in a data driven manner and is less susceptible to getting trapped in local minima. On the other hand, Posest performs a gradient-descent based error minimization in the vicinity of the solution with which it is initialized. Therefore it is easier to get trapped in local minima, in case of error functions that are not smooth and unimodal. The error in the correspondences in the *N* set affect the smoothness of the error function and leads Posest to suboptimal solutions compared to RANSAC.

The above findings are confirmed also by the second experiment which analyzes the performance of the considered initialization methods on the FIRE dataset. Table III and Fig. 5 show that the performance of the two methods is similar, with the largest difference being in A. As presented earlier, this subset contains image pairs acquired at temporally displaced examinations. They feature visual anatomical differences due to the progression or remission of retinopathy. This affects the positioning of the image features over the eye surface, and

consequently, the smoothness of the error function. Therefore, RANSAC outperforms Posest.

Given these results, RANSAC was selected to provide an initial solution to the PSO-based optimization technique.

## F. Search Space Variant Comparison

This experiment has the objective of identifying the most adequate RANSAC and PSO combination for estimating the solution **S** to the registration. The variants described in Section III-I are compared. Variants C, R-C and R-F run p/g of 10 k/300 in a single PSO stage. Variants CF and R-CF run 10 k/150 in each of their two PSO stages. Thus, all variants, except R, have the same budget.

Table IV shows that among variants without RANSAC initialization (C and CF), variant CF offers the best result. It demonstrates that performing an initial coarse search to obtain a rough estimate and refining it afterwards provides the most accurate results. Standalone RANSAC (R) is shown to underperform in almost every condition.

Variants that utilize RANSAC initialization (R-C, R-F and R-CF) outperform their counterparts without initialization. In general, R-F outperforms its competitors, with the largest difference being in A. This experiment shows that with a good



Fig. 7. Registration success for for the four eye model variants. The x axis marks, in pixels, the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.

#### TABLE IV

PERCENTAGE OF AUC FROM FIG. 6 IN EACH FIRE CATEGORY FOR THE 6 VARIANTS FOR SEARCH SPACE EXPLORATION. BOLD SHOWS THE HIGHEST SCORE

Variant	S	$\mathcal{P}$	$ $ $\mathcal{A}$	FIRE
С	0.951	0.324	0.597	0.682
CF	0.952	0.453	0.537	0.724
R	0.933	0.209	0.549	0.624
R-C	0.960	0.532	0.603	0.765
R-CF	0.955	0.516	0.543	0.750
R-F	0.958	0.541	0.660	0.773

#### TABLE V

PERCENTAGE OF AUC FROM FIG. 7 IN EACH FIRE CATEGORY FOR THE FOUR EYE MODEL VARIANTS. BOLD SHOWS THE HIGHEST SCORE

Model	S	$\mathcal{P}$	$\mathcal{A}$	FIRE
Plane	0.926	0.164	0.574	0.608
Sphere	0.945	0.385	0.617	0.703
FO ellipsoid	0.951	0.498	0.626	0.750
Ellipsoid	0.958	0.541	0.660	0.773

initialization, focusing the budget around the initial estimation provides the most accurate solution.

#### G. Eye Model Variant Comparison

This experiment compares the four variants proposed in Section III-J so as to identify which performs best in terms of registration accuracy. Results are shown in Table V and Fig. 7. In all cases, the ellipsoid is shown to outperform the other variants in the majority of image pairs. It is shown that the registration accuracy increases along with the degrees of freedom. This is the case for all three FIRE subsets and it is attributed to the better approximation of the retina's shape and pose. This experiment proves that independently from the characteristics of an image pair, a model that is able to better approximate the actual shape of the eye assists in performing more accurate registration.

# H. Impact of PSO Budget

We investigate how key PSO parameters impact the accuracy and performance of the RIR method. Different combinations of p/g can provide the same computational budget, but the solutions they reach may differ considerably. The performance of budgets ranging from 1000 to 9 million particles per swarm is studied. Each of these budgets is distributed across 200, 250, 300, 350 and 400 generations.

The results of these experiments are shown in Fig. 8. Note that the x axis is logarithmic. For a given budget, distributing it across 300 generations provides the most accurate results. For budgets larger than 3 million particles, the framework reaches asymptotic results, meaning that an increase in budget, barely improves the registration accuracy. Given these results, a budget p/g of 10 k/300 appears to be the best choice. The chosen budget is high compared to budgets used in other domains [59], [61]. This is due to the fact that retinal images provide a very limited view of the surface of the retina. Variations on the shape of the eye model have a large impact towards the edges of the image and small towards the center. As such, large variations in the locations of the particles may yield small differences on the observed retinal surface, requiring a high amount of particles to perform an accurate estimation.

# I. Multiple Swarms

This experiment evaluates the effect of executing the RANSAC and PSO processes several parallel times, to palliate the non-deterministic effect of the registration framework. The framework is executed 10 times, with each execution having a different s ranging from 1 to 10 swarms.

Table VI shows the AUC for every execution when registering the FIRE dataset. Fig. 9 shows the result for the execution with odd s, for clarity purposes. The results show that while the registration framework is non-deterministic, the variability in the registration accuracy when executing the framework with diverse s is small. For the registration method in this work, s = 3is selected, as it provides the best robustness/computational performance trade-off.

# J. Comparison With State of the Art RIR Methods

We compared the accuracy of the registration method with several other existing approaches. The list of methods is shown



Fig. 8. Budget study on registration success. y axis represents AUC for a given budget. Note that while the full range for y axis is from 0 to 1, here it is zoomed in. x axis indicates the total budget used in the registration, in logarithmic scale.



Fig. 9. Registration success for an increasing amount of swarms. The x axis marks, in pixels, the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.

TABLE VI PERCENTAGE OF AREA UNDER THE CURVE (AUC) IN FIG. 9 IN EACH FIRE CATEGORY FOR AN INCREASING NUMBER OF INDEPENDENT SWARMS. BOLD SHOWS THE HIGHEST SCORE

Swarms	S	$\mathcal{P}$	$ \mathcal{A} $	FIRE
1	0.945	0.370	0.623	0.699
2	0.945	0.383	0.623	0.703
3	0.945	0.383	0.634	0.705
4	0.945	0.377	0.640	0.703
5	0.931	0.374	0.634	0.693
6	0.944	0.378	0.614	0.700
7	0.944	0.378	0.614	0.700
8	0.945	0.380	0.620	0.702
9	0.945	0.378	0.623	0.702
10	0.944	0.381	0.643	0.704

TABLE VII COMPETING REGISTRATION METHODS USED FOR COMPARISON, IN CHRONOLOGICAL ORDER

Method	Author	Year	Reference
SIFT	Lowe	2004	[52]
SURF	Bay et al.	2006	[53]
GDB-ICP	Yang et al.	2007	[41]
Harris-PIIFD	Chen et al.	2010	[33]
ED-DB-ICP	Tsai et al.	2010	[42]
RIR-BS	Chen et al.	2011	[64]
WGTM	Izadi and Saeedi	2012	[65]
ATS-RGM	Serradell et al.	2015	[66]
EyeSLAM	Braun et al.	2018	[67]
GFEMR	Wang et al.	2019	[68]
VOTUS	Motta et al.	2019	[63]

in Table VII. Experiments for Generalized Dual Bootstrap-Iterative Closest Point (GDB-ICP) [41] and Harris-PIIFD framework [33]. have been performed by the authors. For GDB-ICP, the C++ implementation provided by the authors<sup>1</sup> was used. For the original Harris-PIIFD framework, a MATLAB implementation was utilized. For the remaining methods, such as Vessel Optimal Transport for fundUS image alignment (VOTUS) and Gaussian Field Estimator with Manifold Regularization (GFEMR) the shown results are taken from [63], where experiments were conducted with the same dataset and methodology as in this manuscript.

Fig. 10 shows the result plot for the experiments performed by the authors, and Table VIII show the AUC for all methods. The proposed framework outperforms most of the competing methods, with the closest competitors being VOTUS and GFEMR. When compared to GFEMR, the proposed method outperforms in S and A, but slightly underperforms in P. The proposed method slightly outperforms VOTUS in S and slightly underperforms in A, but clearly underperforms in P. Overall, the proposed method is closer to the current leading method than to its trailing one.

#### K. Execution Time

This experiment aims to compare the proposed framework with GDB-ICP [41] and the original Harris-PIIFD framework [33] in terms of execution time. It also studies the execution

<sup>1</sup>http://www.vision.cs.rpi.edu/gdbicp/exec/



Fig. 10. Registration success for the proposed method, Harris-PIIFD and GDB-ICP. The x axis marks, in pixels, the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.

#### TABLE VIII

PERCENTAGE OF AUC IN EACH FIRE CATEGORY FOR THE PROPOSED METHOD, AND OTHER STATE-OF-THE-ART METHODS. BOLD SHOWS THE HIGHEST SCORE, ITALICS SHOW THE SECOND BEST RESULT. RESULTS FOR THE PROPOSED METHOD, HARRIS-PIIFD AND GDB-ICP CALCULATED BY THE AUTHORS FROM FIG. 10, REST ARE EXTRACTED FROM [63]

Method	S	$\mathcal{P}$	$\mathcal{A}$	FIRE
Proposed	0.958	0.542	0.660	0.773
VOTUS	0.934	0.672	0.681	0.812
GFEMR	0.812	0.607	0.474	0.702
SIFT+WGTM	0.837	0.544	0.407	0.685
GDB-ICP	0.814	0.303	0.303	0.576
Harris-PIIFD	0.900	0.090	0.443	0.553
ED-DB-ICP	0.604	0.441	0.497	0.533
SURF+WGTM	0.835	0.061	0.069	0.472
RIR-BS	0.772	0.049	0.124	0.440
EyeSLAM	0.308	0.224	0.269	0.273
ATS-RGM	0.369	0.000	0.147	0.211

TABLE IX AVERAGE REGISTRATION TIME PER IMAGE PAIR, IN SECONDS, FOR THE PROPOSED METHOD, HARRIS-PIIFD AND GDB-ICP

	Proposed	Harris-PIIFD	GDB-ICP
Time (s)	198	11	73

times of diverse parts of the proposed algorithm to identify potential bottlenecks.

Table IX shows average registration time per image pair, in seconds, for the proposed method, Harris-PIIFD and GDB-ICP. While Harris-PIIFD is the fastest performing method of the three, in Section IV-J it was shown that its accuracy is poor compared to competing methods. Second fastest method is GDB-ICP. Finally, the slowest method is the proposed one. The proposed method has been timed with the PSO stage running on GPU, an option that is not available in the used implementations of the competing methods.

Table X shows average time required for each of the stages of the proposed method to perform registration for an image pair. The fastest stage is RANSAC, with each execution taking less than 1 second, followed by the PSO process, which has been implemented utilizing GPU acceleration. There exist apparent bottlenecks in the keypoint extraction and matching, and the

TABLE X AVERAGE PROCESS TIME FOR DIVERSE STAGES OF THE PROPOSED METHOD

	Keypoint extraction and matching	Single RANSAC	Single PSO	Image formation
Time (s)	80	> 1	25	40

image formation stages, for which no parallelization has been implemented. Keypoint extraction is particularly slow due to the bifurcation extraction, as it requires to perform vessel segmentation. Image formation is slow due to the tracing of image points of  $F_t$  from the test camera location to the eye model surface  $\mathcal{E}$ , and their projection back to the reference camera location.

While the proposed method is slower than competing methods, it provides more accurate registration and the execution time is not prohibitively large for being used in an examination session. Bottlenecks on the bifurcation extraction, as well as the image formation, have been identified, and will be addressed in future work.

## L. Qualitative Results

In this section, qualitative results of REMPE in the 3 categories of the FIRE dataset are presented. Fig. 11 shows registration results for 6 image pairs of the FIRE [62] dataset. Image pairs from the top row correspond to category S, so they contain no anatomical differences and have large overlapping surface. Image pairs from the middle row are extracted from  $\mathcal{P}$ , so they contain no anatomical differences, but present small overlap. Image pairs from the bottom row belong to category  $\mathcal{A}$ , so there exist anatomical differences within each image pair, but the overlap is large.

These results show that the proposed RIR method performs accurate registration for image pairs for varying degrees of overlap, and with or without anatomical differences. This is an important tool for clinicians, as it may facilitate visual analysis of image pairs to detect retinal changes such as early arteriolosclerosis, which is important for early diagnosis of hypertension.



Fig. 11. Registration results for the FIRE dataset. Top row shows registration for image pairs from category  $\mathcal{S}$ , mid row for category  $\mathcal{P}$  and bottom row for category  $\mathcal{A}$ . Original registered images had a resolution of  $2912 \times 2912$ . Results displayed in a checkerboard mosaic, alternating patches from both images. Images in the figure have high resolution, allowing to zoom in to study them in detail.

#### V. CONCLUSION

A RIR framework for fundus images is proposed. The framework, which is publicly available,<sup>2</sup> is based on the simultaneous estimation of the relative 3D pose of the cameras that acquired the retinal images to be registered and the parameters and the orientation of an ellipsoidal model of the eye. To perform this, we rely on SIFT and bifurcation matches, as combining these types of features provides increased accuracy. Keypoints are employed with a spherical eye model to perform an initial camera pose estimation using RANSAC, which proves to provide more accurate results than Posest. The pose is subsequently refined by utilizing PSO with an ellipsoidal model of the eye for which the parameters are also estimated. Utilizing an accurate eye model for the registration allows for more accurate registration than alternative polynomial transformations.

The performed experimental evaluation validates the proposed method using the publicly available<sup>3</sup> FIRE dataset. Regarding eye models, an ellipsoid, which has the potential to approximate the actual shape of the eye more accurately than alternative models such as a sphere or a plane is also shown to improve accuracy. Finally, experiments show the proposed method to be more accurate and robust than most of the stateof-the-art methods it was compared with, while being able to compete head to head with VOTUS, which is currently the best performing method.

Two important elements of this work is the introduction of the world model and the investigation of an alternative initialization step. The world model is the cornerstone of the geometrical part of the proposed method, upon which the rest of the techniques build to achieve retinal image registration. The consideration of Posest as an alternative initialization step, while conceptually promising, did not give the expected results during its experimental evaluation. Therefore, its selection over RANSAC is not justified. Finally, by making available the code implementing the REMPE method to the scientific community, we facilitate further experimentation and its application in real environments.

#### REFERENCES

- M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, Jan. 2010.
- [2] A. Grosso *et al.*, "Hypertensive retinopathy revisited: Some answers, more questions," *Brit. J. Ophthalmol.*, vol. 89, no. 12, pp. 1646–1654, 2005.
- [3] R. P. Danis and M. D. Davis, *Proliferative Diabetic Retinopathy*. Totowa, NJ, USA: Humana Press, 2008, pp. 29–65.
- [4] C. Hernandez-Matas, A. A. Argyros, and X. Zabulis, "Retinal image preprocessing, enhancement, and registration," in *Computational Retinal Image Analysis*, E. Trucco, T. MacGillivray, and Y. Xu, Eds. New York, NY, USA: Academic, 2019, pp. 59–77.
- [5] N. Meitav *et al.*, "Improving retinal image resolution with iterative weighted shift-and-add," *J. Opt. Soc. America A*, vol. 28, no. 7, pp. 1395–1402, 2011.
- [6] G. Molodij *et al.*, "Enhancing retinal images by extracting structural information," *Opt. Commun.*, vol. 313, pp. 321–328, 2014.
- [7] C. Hernandez-Matas and X. Zabulis, "Super resolution for fundoscopy based on 3D image registration," in *Proc. IEEE Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, Aug. 2014, pp. 6332–6338.
- [8] L. Hubbard et al., "Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study," *Ophthalmology*, vol. 106, pp. 2269–80, 1999.
- [9] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A featurebased technique for joint, linear estimation of high-order image-to-mosaic transformations: Mosaicing the curved human retina," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 412–419, Mar. 2002.
- [10] N. Ryan, C. Heneghan, and P. de Chazal, "Registration of digital retinal images using landmark correspondence by expectation maximization," *Image Vision Comput.*, vol. 22, no. 11, pp. 883–898, 2004.
- [11] P. C. Cattin *et al.*, *Retina Mosaicing Using Local Features*. Berlin Germany: Springer, 2006, pp. 185–192.
- [12] H. Narasimha-Iyer, A. Can, B. Roysam, H. L. Tanenbaum, and A. Majerovics, "Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1436–45, Aug. 2007.
- [13] G. Troglio et al., Unsupervised Change-Detection in Retinal Images by a Multiple-Classifier Approach. Berlin, Germany: Springer, 2010, pp. 94–103.
- [14] 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. IEEE 37th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, Aug. 2015, pp. 5650–5654.
- [15] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. A. Argyros, "Retinal image registration under the assumption of a spherical eye," *Computerized Med. Imag. Graph.*, vol. 55, pp. 95–105, 2017.
- [16] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration through simultaneous camera pose and eye shape estimation," in *Proc. IEEE Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, Aug. 2016, pp. 3247–3251.

<sup>&</sup>lt;sup>2</sup>[Online]. Available: http://www.ics.forth.gr/cvrl/rempe

<sup>&</sup>lt;sup>3</sup>[Online]. Available: http://www.ics.forth.gr/cvrl/fire

- [17] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "An experimental evaluation of the accuracy of keypoints-based retinal image registration," in *Proc. IEEE Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, Jul. 2017, pp. 377–381.
- [18] A. V. Cideciyan *et al.*, "Registration of high resolution images of the retina," *SPIE Med. Imag.*, vol. 1652, pp. 310–322, Jun. 1992.
- [19] R. Kolar, V. Harabis, and J. Odstrcilik, "Hybrid retinal image registration using phase correlation," *Imag. Sci. J.*, vol. 61, no. 4, pp. 369–384, 2013.
- [20] E. Peli, R. A. Augliere, and G. T. Timberlake, "Feature-based registration of retinal images," *IEEE Trans. Med. Imag.*, vol. 6, no. 3, pp. 272–278, Sep. 1987.
- [21] G. K. Matsopoulos, P. A. Asvestas, N. A. Mouravliansky, and K. K. Delibasis, "Multimodal registration of retinal images using self organizing maps," *IEEE Trans. Med. Imag.*, vol. 23, no. 12, pp. 1557–1563, Dec. 2004.
- [22] S. Lee *et al.*, "Retinal image mosaicing using the radial distortion correction model," *Proc. SPIE*, vol. 6914, 2008, pp. 6914–6914–9.
- [23] T. E. Choe *et al.*, "2-D registration and 3-D shape inference of the retinal fundus from fluorescein images," *Med. Image Anal.*, vol. 12, no. 2, pp. 174–190, 2008.
- [24] A. R. Chaudhry and J. C. Klein, "Ophthalmologic image registration based on shape-context: Application to fundus autofluorescence (FAF) images," *VIIP*, pp. 1–7, 2008.
- [25] S. Gharabaghi et al., "Retinal image registration using geometrical features," J. Digit. Imag., vol. 26, no. 2, pp. 248–258, 2013.
- [26] P. S. Reel *et al.*, "Robust retinal image registration using expectation maximisation with mutual information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 1118–1122.
- [27] L. Chen, X. Huang, and J. Tian, "Retinal image registration using topological vascular tree segmentation and bifurcation structures," *Biomed. Signal Process. Control*, vol. 16, pp. 22–31, 2015.
- [28] J. A. Lee *et al.*, "A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 1046–1053.
- [29] C. Liu *et al.*, "Retinal image registration via feature-guided gaussian mixture model," *J. Opt. Soc. America A*, vol. 33, no. 7, pp. 1267–1276, Jul. 2016.
- [30] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inf. Technol. Biomedicine*, vol. 3, no. 1, pp. 47–60, Mar. 1999.
- [31] Y. Lin and G. Medioni, "Retinal image registration from 2D to 3D," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., Jun. 2008, pp. 1–8.
- [32] K. M. Adal *et al.*, "A hierarchical coarse-to-fine approach for fundus image registration," in *Proc. Int. Workshop Biomed. Image Registration*, 2014, vol. 8545, pp. 93–102.
- [33] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707–1718, Jul. 2010.
- [34] Z. Ghassabi et al., "An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors," EURASIP J. Image Video Process., vol. 2013, no. 1, p. 25, 2013.
- [35] G. Wang *et al.*, "Robust point matching method for multimodal retinal image registration," *Biomed. Signal Process. Control*, vol. 19, pp. 68–76, 2015.
- [36] S. K. Saha et al., "A two-step approach for longitudinal registration of retinal images," J. Med. Syst., vol. 40, no. 12, p. 277, 2016.
- [37] F. Laliberte, L. Gagnon, and Yunlong Sheng, "Registration and fusion of retinal images-an evaluation study," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 661–673, May 2003.
- [38] K. Deng *et al.*, "Retinal fundus image registration via vascular structure graph matching," *J. Biomed. Imag.*, vol. 2010, pp. 14:1–14:13, Jan. 2010.
  [39] A. Perez-Rovira, R. Cabido, E. Trucco, S. J. McKenna, and J. P.
- [39] A. Perez-Rovira, R. Cabido, E. Trucco, S. J. McKenna, and J. P. Hubschman, "RERBEE: Robust efficient registration via bifurcations and elongated elements applied to retinal fluorescein angiogram sequences," *IEEE Trans. Med. Imag.*, vol. 31, no. 1, pp. 140–150, Jan. 2012.
- [40] 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.
- [41] G. Yang, C. V. Stewart, M. Sofka, and C. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1973–1989, Nov. 2007.
- [42] C. L. Tsai, C. Li, G. Yang, and K. 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, 2010.

- [43] Z. Hossein-Nejad and M. Nasri, "A-RANSAC: Adaptive random sample consensus method in multimodal retinal image registration," *Biomed. Signal Process. Control*, vol. 45, pp. 325–338, 2018.
- [44] R. Ramli et al., "Feature-based retinal image registration using D-saddle feature," J. Healthcare Eng., vol. 2017, 2017, Art. no. 1489524.
- [45] G. Noyel *et al.*, "Registration of retinal images from Public Health by minimising an error between vessels using an affine model with radial distortions," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, Apr. 2019, pp. 561–564.
- [46] E. Sabanovic, G. Stankevièius, and D. Matuzevièius, "Deep neural network-based feature descriptor for retinal image registration," in *Proc. IEEE 6th Workshop Adv. Inf., Electron. Elect. Eng.*, Nov. 2018, pp. 1–4.
- [47] T. Che et al., "Deep group-wise registration for multi-spectral images from fundus images," *IEEE Access*, vol. 7, pp. 27 650–27 661, 2019.
- [48] F. Zhang et al., "Automated quality classification of colour fundus images based on a modified residual dense block network," *Signal, Image Video Process.*, vol. 14, pp. 215–223, Aug. 2019.
- [49] J. A. Lee *et al.*, "A deep step pattern representation for multimodal retinal image registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, 2019, pp. 5076–5085.
- [50] R. Navarro *et al.*, "Accommodation-dependent model of the human eye with aspherics," *J. Opt. Soc. America A*, vol. 2, no. 8, pp. 1273–1280, Aug. 1985.
- [51] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [52] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [53] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Comput. Vision Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [54] T. Chanwimaluang, "Advanced retinal imaging: Feature extraction, 2-D registration, and 3-D reconstruction," Ph.D. dissertation, Oklahoma State Univ, Stillwater, OK, USA, 2006.
- [55] M. A. Fischler *et al.*, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [56] M. Lourakis and X. Zabulis, Model-Based Pose Estimation for Rigid Objects. Berlin, Germany: Springer, 2013, pp. 83–92.
- [57] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," J. Soc. Ind. Appl. Math., vol. 11, no. 2, pp. 431–441, 1963.
- [58] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, 2007.
- [59] I. Oikonomidis, N. Kyriazis, and A. A. Argyros, "Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints," in *Proc. IEEE Int. Conf. Comput. Vision*, Nov. 2011, pp. 2088–2095.
- [60] N. Kyriazis and A. Argyros, "Scalable 3D tracking of multiple interacting objects," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2014, pp. 3430–3437.
- [61] X. Zabulis, M. Lourakis, and P. Koutlemanis, "3D object pose refinement in range images," in *Proc. Int. Conf. Comput. Vision Syst.*, 2015, pp. 263–274.
- [62] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. A. Argyros, "Fire: Fundus image registration dataset," *J. Model. Ophthalmol.*, vol. 1, no. 4, pp. 16–28, 2017.
- [63] D. W. Casaca, and A. Paiva, "Vessel optimal transport for automated alignment of retinal fundus images," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6154–6168, Dec. 2019.
- [64] L. Chen, Y. Xiang, Y. Chen, and X. Zhang, "Retinal image registration using bifurcation structures," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2169–2172.
- [65] M. Izadi and P. Saeedi, "Robust weighted graph transformation matching for rigid and nonrigid image registration," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4369–4382, Oct. 2012.
- [66] E. Serradell, M. A. Pinheiro, R. Sznitman, J. Kybic, F. Moreno-Noguer, and P. Fua, "Non-rigid graph registration using active testing search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 625–638, Mar. 2015.
- [67] D. Braun et al., "EyeSLAM: Real-time simultaneous localization and mapping of retinal vessels during intraocular microsurgery," Int. J. Med. Robot. Comput. Assited Surgery, vol. 14, no. 1, 2018, Art. no. e1848.
- [68] J. Wang *et al.*, "Gaussian field estimator with manifold regularization for retinal image registration," *Signal Process.*, vol. 157, pp. 225–235, 2019.