Retinal Image Registration Based on Keypoint Correspondences, Spherical Eye Modeling and Camera Pose Estimation

Carlos Hernandez-Matas[1,2], Xenophon Zabulis[1] and Antonis A. Argyros[1,2]

Abstract—In this work, an image registration method for two retinal images is proposed. The proposed method utilizes keypoint correspondences and assumes a spherical model of the eye. Image registration is treated as a pose estimation problem, which requires estimation of the rigid transformation that relates the two images. Using this estimate, one image can be warped so that it is registered to the coordinate frame of the other. Experimental evaluation shows improved accuracy over state-of-the-art approaches as well as robustness to noise and spurious keypoint correspondences. Experiments also indicate the method's applicability to diagnostic image enhancement and comparative analysis of images from different examinations.

I. INTRODUCTION

Small vessel structure and function assessment can lead to more accurate and timely diagnosis of diseases whose common denominator is vasculopathy, i.e. hypertension and diabetes [1]. Small vessels exist in all internal and external organs. Of them, the retina provides an open and accessible window for assessing their condition. Retinal vessels are imaged through fundoscopy, an efficient and non-invasive imaging technique that is suitable for screening. Accurate image registration is of interest in the comparison of images from different examinations [2] and in the combination of multiple images into larger [3] or enhanced [4] ones.

Image registration has been employed frequently on slightly overlapping images of the same examination, to create mosaic images of large tissue areas, i.e. [3]. Small overlap increases examination efficiency, but increases registration difficulty as it is based on less data. This difficulty is tackled by strong registration cues, such as keypoint correspondences, i.e. [5].

Not frequently, image registration has been employed to register images of (approximately) the same retinal region. Motivation is twofold. First, to combine images from the same examination into an image of higher resolution, facilitating more precise measurements, [6], [7], [4]. Second, to register images from different examinations and comparatively analyze them [2], [8].

In this work, the image registration problem refers to a pair of images, the reference and the test image. Its solution is the aligning transformation that warps the test image so that the same physical points are imaged in the same pixel coordinates as in the reference image. Henceforth, image registration methods which provide a solution by means of transformation(s) upon the image plane are characterized as "2D", while methods which account for the retina as a surface imaged from different views as "3D".

The proposed method focuses on the cue to image registration due to keypoint correspondences. The additional value of other cues is acknowledged; i.e. edge, bifurcation matching. The proposed framework is open to additional cues and their adoption is left for future work.

II. RELATED WORK

For retinal image registration, overlapping image regions have been matched using similarity of intensities over spatial regions [9] or the frequency domain [10], keypoint feature correspondences [5], retinal feature matching i.e. vessel trees [11], bifurcations [12]. Feature-based approaches are preferred in 3D approaches, as point correspondences comprise a relatively stronger cue for estimating the motion between two images and, also, are robust to local image differences.

Retinal image registration has been studied using 2D and 3D transformation models. 2D models do not explicitly account for perspectivity [11], though some [12] employ non-linear transformations for this purpose. 3D models account for perspectivity, but require the shape of the imaged surface. Consideration of perspectivity improves image registration. Even simple surface models, as a planar patch, were shown to promote registration accuracy [4]. At the other end, in [5], the retinal surface is reconstructed to achieve registration. However it requires a stereo reconstruction of the retina, which for significantly overlapping images is inaccurate due to the very short baseline.

Fundus imaging has been modeled by the pinhole camera model [5]. Usually lens distortion has been judged as negligible, due to the fine optics of fundus cameras. Visual distortions due to the cornea, the eye lens, the vitreous humor, as well as pulsation, have been approximated as negligible. We also follow these approximations, acknowledging that compensating for pertinent distortions, would increase the accuracy of the proposed method.

The proposed method utilizes a 3D cost optimization method that is robust to correspondence errors and copes with local minima. Efficiency is supported by a parallel implementation and evaluation shows improved performance with respect to state-of-the-art. The method is open to the addition of more visual cues (i.e. due to edges, intensity).

III. METHOD

The proposed method estimates the rigid transformation $\{R, \mathbf{t}\}$ that relates the reference (F_0) and the test (F_r) image,

[1]Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece.

[2]Computer Science Department, University of Crete, Heraklion, Greece. {carlos, zabulis, argyros} at ics.forth.gr
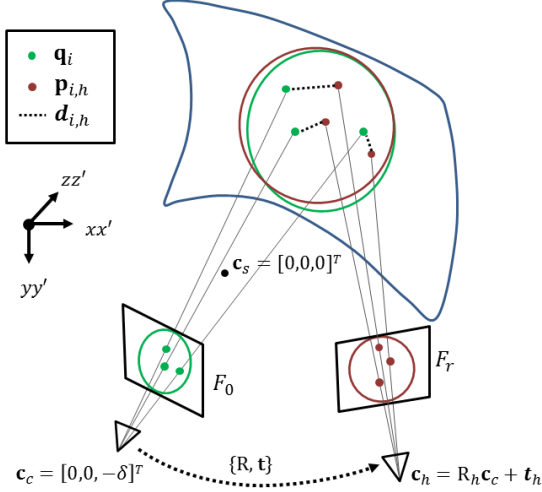
Fig. 1. The geometry of the objective function.

as the motion between the two frames. It assumes a spherical eye model centered at $\mathbf{c}_s = [0,0,0]^T$ and, at distance δ, a calibrated camera, located at $\mathbf{c}_c = [0,0,-\delta]^T$; K is the intrinsic camera matrix. Equivalently, this estimate can be also expressed as the relative pose of the retina between the two frames if a static camera is assumed. Hypothesis, or "candidate pose", with id h regards camera motion $\{R_h, \mathbf{t}_h\}$.

Hypotheses are generated in a 6D search space around an initial pose estimate. This estimate is optional and, if unavailable, is the identity transformation $\{I, \mathbf{0}\}$.

Candidate poses $\{R_h, \mathbf{t}_h\}$ are generated and evaluated with the aid of objective function $o(\{R_h, \mathbf{t}_h\})$ as follows. Given the eye's model, an image point can be ray-traced to the 3D location on the retina that it images. Ideally, in the correct pose, corresponding keypoints should co-occur, while for increasingly inaccurate poses their distances would accordingly increase (see Fig. 1). Function $o(\cdot)$ accumulates these spatial disparities exhibiting a minimum (ideally 0) at the correct configuration. As correspondences across the two images may include matching errors, function $o(\cdot)$ uses a percentile of the matches to safeguard from outliers.

The 6D, iterative, search for the pose that yields the best score is optimized using Particle Swarm Optimization (PSO) [13], a derivative-free stochastic optimization framework .

A. Keypoint Features

SURF features [14] were selected as keypoints, due to their improved robustness to spurious matches and their localization accuracy [15]. SURF keypoints have been reported to be more reliable than other keypoints in retinal images [4]. Keypoints are matched as in [16]. For color images, the "green" channel is used, as it offers a higher contrast [17].

B. Objective Function

The objective function returns a score for candidate pose $\{R_h, \mathbf{t}_h\}$. Given the (reference or candidate) camera pose, keypoints can be ray-traced to 3D locations on the spherical model of the retina. This is achieved by finding its intersection with the optical ray through the keypoint. The Navarro eye model [18], for the eye dimensions ($\rho = 12\,mm$) is utilized, with the origin at the center of this spherical eye model. The sphere equation is: $\|\mathbf{x} - \mathbf{c}_s\|^2 = \rho^2$ where $\mathbf{x}$ is a point on the sphere, $\mathbf{c}_s$ is the center of the sphere and ρ its radius. To find $\mathbf{x}$, the ray equation from the camera center $\mathbf{c}$ ($\mathbf{c}_c$ or $\mathbf{c}_h$) and through pixel $\mathbf{u}$ is solved for λ:

$$\mathbf{x} = P^+\mathbf{u} + \lambda\mathbf{c}, \qquad (1)$$

where $P^+ = P^T(PP^T)^{-1}$ (see Eq. 6.13, [19, p. 162]).

Let $\mathbf{q}_i$ the 3D locations of keypoints in F_0. The 3D locations of keypoints from the test image are determined by pose $\{R_h, \mathbf{t}_h\}$ (see Fig. 1). Let $\mathbf{p}_{i,h}$ the 3D locations of keypoints in F_r, for pose hypothesis h. The 3D distances of corresponding 3D locations are $d_{i,h} = |\mathbf{q}_i - \mathbf{p}_{i,h}|$. A selection for $o(\cdot)$ could be the accumulation of distances $d_{i,h}$. To obtain robustness to spurious matches, a percentile of these matches is used instead:

$$o(\{R_h, \mathbf{t}_h\}) = \sum_j d_{j,h}, \qquad (2)$$

where j enumerates the smallest 70% values of $d_{i,h}$. To combine more cues Eq. 2 can be amended with further terms, i.e. accumulating edge distances.

C. Particle Swarm Optimization

Minimization of $o(\cdot)$ is performed via PSO, an iterative, stochastic optimization method. PSO involves n_p particles that evolve for n_g generations, amounting to a total of $n_p \cdot n_g$ objective function evaluations. This product is called the "budget" of an optimization and represents a trade-off between accuracy and speed of execution. The particles in the first generation are initialized randomly in the search space, communicate in every generation, and converge towards particles that yield the best values of $o(\cdot)$. Due to the independence of particles pertaining to each generation, their execution is performed in parallel on the CPU.

The search space is a 6D hypercube, centered around an initial pose. In the general case, this pose is the same with the reference pose $\{I, \mathbf{0}\}$. Poses are parameterized as translations along and rotations about the 3 axes. Thus $\mathbf{t}_h = [t_{x,h}\, t_{y,h}\, t_{z,h}]^T$, while $R_h = R_x(r_{\theta,h}) \cdot R_y(r_{\phi,h}) \cdot R_z(r_{\omega,h})$. The search space around the initial pose is denoted as $[\mu_x, \mu_y, \mu_z, \gamma_\theta, \gamma_\phi, \gamma_\omega]$, meaning that $t_{x,h} \in [-\mu_x/2, \mu_x/2]$ and correspondingly for the rest of dimensions.

D. Cascaded Particle Swarm Optimization

Despite the robustness to local minima, due to multiple particles that simultaneously search pose space, particle communication in PSO may still be entrapped in a local minimum. Though local minima do not provide a better score than the correct solution, this phenomenon can occur due

to sparse particle distribution. To cope with these cases, 3 cascaded variants (B, C, and D) of the original optimization process (A) were considered. These variants perform chained PSO executions. In the first stage, the initial pose is $\{I, \mathbf{0}\}$. The search range is a 6D hypercube around this pose. In succeeding stages, the result of the preceding execution is pipelined as the initial pose estimate.

In each execution, 64 particles are reserved to be initialized at the corners of the hypercube, to ensure a widespread particle allocation. In succeeding executions of the optimization process, an extra particle is initialized at the estimate of the previous stage, so that it is considered in the optimization. The rest of the particles are randomly initialized.

A. Single stage. This is the baseline method of Sec. III-C.

B. Coarse-to-fine. PSO is executed twice. In the second execution, the search space is a smaller hypercube centered at the estimate of the first execution. The same budget is retained, but the hypercube is constrained in all 6 dimensions thus defining a denser search. In the second execution, the range is $[\frac{\mu_x}{\alpha}, \frac{\mu_y}{\alpha}, \frac{\mu_z}{\alpha}, \frac{\gamma_\theta}{\alpha}, \frac{\gamma_\phi}{\alpha}, \frac{\gamma_\omega}{\alpha}]$ and $\alpha = 3$.

C. Pairing of conflicting dimensions. PSO is executed thrice. It is based on the observation that in the 6D neighborhood of the correct pose local minima occur. The 6D locations of these minima correspond to poses for which points $\mathbf{p}_{i,h}$ occur at very proximate locations. Small changes in r_ϕ and in t_x can have a similar effect on $\mathbf{p}_{i,h}$, as they induce displacement vectors that are very similar, thus, different locations in the pose space yield similar objective function scores. A similar effect occurs for t_y and r_θ. Thus, in the second execution, the range is $[\frac{\mu_x}{\beta}, \frac{\mu_y}{\alpha}, \frac{\mu_z}{\beta}, \frac{\gamma_\theta}{\alpha}, \frac{\gamma_\phi}{\beta}, \frac{\gamma_\omega}{\beta}]$, centered around the first estimate. In the third execution, the range is $[\frac{\mu_x}{\alpha}, \frac{\mu_y}{\beta}, \frac{\mu_z}{\beta}, \frac{\gamma_\theta}{\beta}, \frac{\gamma_\phi}{\alpha}, \frac{\gamma_\omega}{\beta}]$ centered around the second estimate. Here $\alpha \ll \beta$, meaning that in the second execution the search will be mostly along μ_y and γ_θ and in the third execution in μ_x and γ_ϕ. Values of $\alpha = 3$, $\beta = 30$ were chosen empirically; future work aims at their optimization.

D. Independent executions. Variant C is executed multiple times (10), independently. The result with the best score is adopted. Motivation stems from countering unfortunate randomization of particle velocities. This differs from a single PSO with 10 times the particles as, due to particle communication, it can be entrapped in a local minimum.

E. Image registration result

Estimate $\{R, \mathbf{t}\}$ is used to warp the test image F_r, so to be registered to the reference F_0. The result, F_w, is parameterized upon the coordinate frame of the reference image F_0. Each pixel coordinate $\mathbf{u}$ in F_0 corresponds to a 3D location $\mathbf{x}$ on the retina (Eq. 1). The intensity at each pixel in F_w is provided by the image formation rule below:

$$F_w(\mathbf{u}) = F_r(P \cdot \mathbf{x}), \qquad (3)$$

where $P = K \cdot [R, \mathbf{t}]$ is the projection matrix of the camera at the estimated pose. For the employed dataset, δ ($\approx 57.7\, mm$) is known from device specifications. Inaccuracies in its value do not affect the registration result, as δ determines only the scale factor up to which the translation component of the

estimated transformation is found; it has no impact on pixel correspondence and, thereby, the formation of F_w.

IV. EXPERIMENTS

Both synthetic and real data were utilized in the experiments. Synthetic data were generated in order to have ground truth about camera poses in evaluation experiments. The data were generated using the texture from a fundus image, to render a retina at multiple, known poses at small and larger departures from the reference pose. The resolution of the simulated camera was 2912×2912 pixels and its Field of View (FOV) was $45°$ in both dimensions. The poses for the synthetic images are denoted $\{t_x, t_y, t_z, r_\theta, r_\phi, r_\omega\}$, where translations are expressed in mm and rotations in degrees.

Real data were acquired from a Nidek AFC-210 fundus camera, with parameters identical to the above. In Fig. 2, 3 real fundus image pairs are shown. Left and middle pairs, were acquired in the same examination and demonstrate a small and a larger transformation, respectively. The images of the right pair, were acquired 1 year apart. In Fig. 3, registration results for these pairs are shown demonstrating registration accuracy. Image registration from temporally different examinations is clinically important, in order to track retinal changes across time and monitor vessel state. In Fig. 2 right, the images present mild changes, as in the width of the arterioles and venules, due to disease. Robustness to pathological alterations is a topic of future work.

Experiments were run on a conventional PC with an i7-4770 CPU, at $3.40\, GHz$ and $16\, GB$ of RAM. The duration of SURF feature matching is $\approx 1.5\, sec$. A single particle evaluation takes less than $0.1\, sec$. Computational time was dominated by PSO which, with variant C, $n_g = 200$ per stage and $n_p = 2500$ per generation being typically about $60\, sec$. For the same parameters, variant D lasts $\approx 600\, sec$.

In evaluation experiments where ground truth was available (Sec. IV-A, IV-B, and IV-C), the accuracy of the proposed method was measured as a pose estimation error. Error is quantified as in [20], which is the average distance between corresponding model points $\mathbf{x}_\kappa$ at the ground truth and the estimated pose. For a ground truth pose $\{R_g, \mathbf{t}_g\}$ and an estimate $\{R_h, \mathbf{t}_h\}$, the error expresses their misalignment and is given by:

$$E = \frac{1}{\nu} \sum_{\kappa=1}^{\nu} |\mathbf{g}_\kappa - \mathbf{e}_\kappa|, \qquad (4)$$

where $\mathbf{g}_\kappa = R_g \cdot \mathbf{x}_\kappa + \mathbf{t}_g$, $\mathbf{e}_\kappa = R_h \cdot \mathbf{x}_\kappa + \mathbf{t}_h$, and ν is the number of model points. Though measuring the error in 3D is more expressive, this error can be converted to 2D error by projecting points $\mathbf{g}_\kappa$, $\mathbf{e}_\kappa$ on F_0 and accumulating the respective 2D distances. The purpose of such a conversion is comparison with a 2D method (in Sec. IV-D).

A. Multiple swarms

This experiment evaluates the effect of distributing budget across multiple swarms has to accuracy, versus assigning its entirety in a single swarm. Fig. 4 shows that running multiple swarms and selecting the one with the best score,
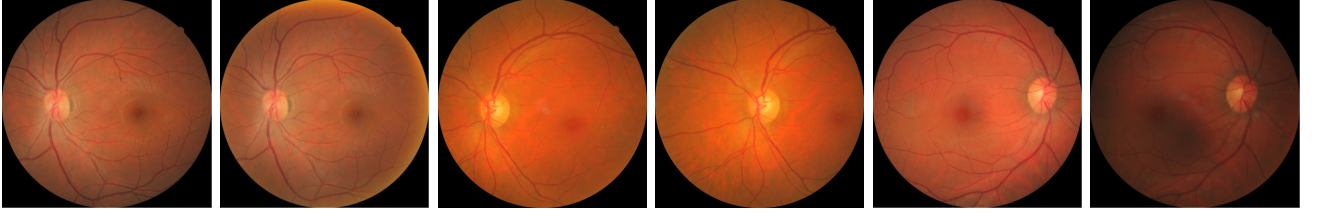
Fig. 2. Fundus images pairs acquired in the same examination (left, middle) and from examinations 1 year apart (right).
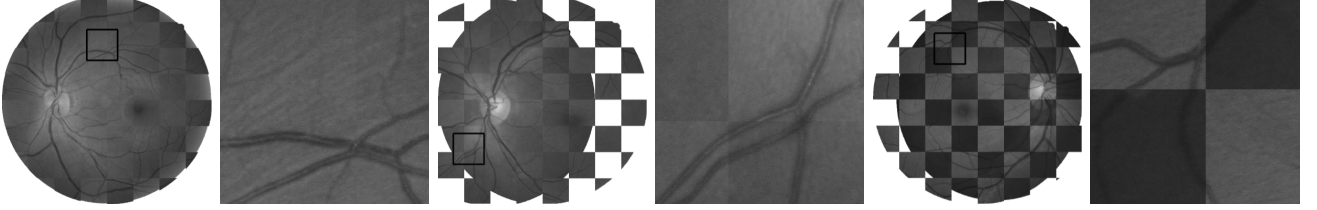

Fig. 3. Registration results for the image pairs of Fig. 2, respectively. The collages show, alternatively, the green channels of the reference and registered image. The marked region with solid black line indicates the magnified image detail shown on the right.
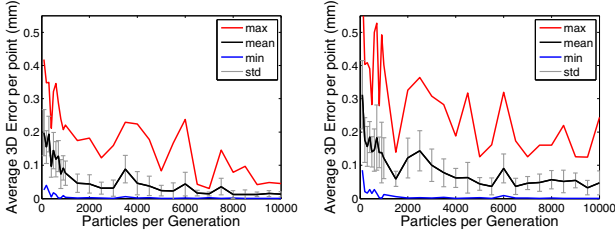

Fig. 4. Registration error evolution with the amount of particles used per generation for our method using 3-stage PSO with 100 generations. Each experiment was run 10 times with 1 PSO swarm.
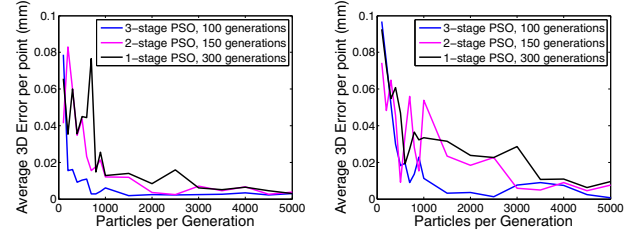
Fig. 5. Registration error evolution vs number of particles for variants A, B, and C using multiple swarms, for small (left) and large transformations (right). Plots at the same horizontal coordinate refer to same budgets.

results in a better estimate than utilizing the entire budget in a single swarm. For example, picking the best result of 10 runs with $n_p = 1000$ is more likely to provide a better result than a single run with $n_p = 10000$. We choose to run 10 independent swarms in each experiment, adopting variant D.

B. Comparison of PSO variants

The PSO variants are evaluated having adopted the strategy of applying multiple swarms (Sec. IV-A). Preliminary experiments indicated that the pairs of conflicting dimensions (t_x-r_ϕ and t_y-r_θ) were causing the optimization to be entrapped in local minima when running variant A. Variants A, B, and C using 10 swarms each (similarly to D) were evaluated with increasing amounts of n_p. Variant A used a 1-stage PSO with $n_g = 300$ generations, B a 2-stage PSO with $n_g = 150$, and C a 3-stage PSO with $n_g = 100$.

The results of this experiment are shown in Fig. 5. Each graph shows the mean registration error for 3 pairs of images with similar magnitude of pose differences; small pose differences are on the left (i,e. $\{0.1, 0.1, 0.3, -0.1, 0.1, 0.1\}$) and larger pose differences on the right (i.e. $\{-1.5, 0, -1, -8, 5, 5\}$). As the center of rotation is $\mathbf{c}_s$, rotations of even $0.1°$ are not negligible (see Fig. 1). It is observed that for an equal budget, the pose estimated by C has the smallest error. This is more pronounced in smaller budgets. Variant C is found to perform better than A and B and adopted.

C. 3D pose estimation accuracy

In this experiment, the proposed method is compared with pose estimation based on point correspondences. Random Sample Consensus (RANSAC) is employed to solve the Perspective-n-Point (PnP) problem [21]. This is a conventional approach to pose estimation when correspondences are available. The approach is robust to spurious matches and estimates object pose given a set of 2D-3D correspondences, through minimization of reprojection error.

After matching keypoints in F_0 and F_r our data are treated as follows, to provide input to RANSAC. Let $\mathbf{u}_i$ and $\mathbf{v}_i$ be the 2D locations of matched keypoints in F_0 and F_r, respectively. By Eq. 1 the 3D locations, $\mathbf{q}_i$, of $\mathbf{u}_i$ are found. Points $\mathbf{q}_i$ and $\mathbf{v}_i$ are provided as input to RANSAC, that estimates the rigid transformation relating the two poses.

Fig. 6 shows that the proposed method, for several combinations of n_p and n_g, provides better pose estimation and thus, better registration, than RANSAC.

Another 3D registration method method is [5], which reconstructs retinal surface given a sufficiently large baseline. Besides this difference, a direct quantitative comparison with this method was unfortunately not possible as the data
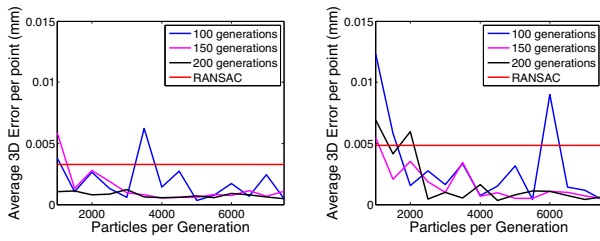
Fig. 6. Registration error for variant D, for 100, 150 and 200 generations vs number of particles, for a small (left) and a large transformation (right) between the 2 images. For RANSAC plot in red, see Sec. IV-C.

Error (pixels)	This work	GDB-ICP [12]
Mean and std all	0.4878 (0.2064)	0.5158 (0.2395)
Mean and std periphery	1.5022 (2.0587)	1.6082 (1.9087)

TABLE I

2D REGISTRATION ERROR (AND STD) FOR ALL CORRESPONDENCES AND FOR THOSE OCCURRING IN IMAGE PERIPHERY.

employed are multimodal, and no ground truth is available for them. By qualitative comparison[1], the registration results from the proposed method exhibit higher accuracy than in the results of [5], in which discontinuities are observed in vessels when comparing registration results as in Fig. 3.

D. 2D registration accuracy

In this experiment, we compare our results, using a budget of $n_g = 200$ and $n_p = 2500$, with Dual Bootstrap Iterative Closest Point [12], a widely employed state-of-the-art 2D retinal image registration method[2]. The 2D mean distances of corresponding points after registration were measured as registration error.

This experiment was performed using 9 real images. Images 2 to 9 were registered to image 1, thus, registering 8 image pairs. A total of 801 control points were found in these image pairs, averaging about 100 points per image pair. Said control points are found using SIFT features, so they are independent from the SURF features employed for the registration. Table I, shows the error for all the control points and for the ones located at the periphery; $|\mathbf{u} - \mathbf{c_i}|^2 > 0.9 \cdot r^2$ where $\mathbf{c_i}$ is the center of the image and r the distance from $\mathbf{c_i}$ to the edge of the image. The proposed method outperforms [12], particularly in periphery of the image, where the effect of employing a clinically correct eye model with a known curvature instead of a quadratic equation is most notable.

V. CONCLUSION

A method for retinal image registration for fundoscopy images was proposed. Experimental testing shows its increased accuracy and robustness compared to state-of-the-art approaches.

The relatively large computational cost of the proposed method is justified by the offline and critical nature of

targeted applications. Acceleration based on parallelization (i.e. using the GPU) is planned for future work.

Further, future work warrants more sources of information for more accurate results and the increase of optimized dimensions to include the shape and size of the retina as well as the camera parameters.

REFERENCES

[1] A. Grosso, F. Veglio, M. Porta, F. Grignolo, and T. Wong, "Hypertensive retinopathy revisited: some answers, more questions." *British Journal of Ophthalmology*, vol. 89, no. 12, pp. 1646–54, 2005.

[2] H. Narasimha-Iyer and A. Can, "Automated analysis of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy," *IEEE Trans. on Biomed. Eng.*, pp. 1084–98, 2006.

[3] A. Can *et al.*, "A feature-based technique for joint linear estimation of high-order image-to-mosaic transformations: Mosaicing the curved human retina." *IEEE PAMI*, vol. 24, no. 3, pp. 412–19, 2002.

[4] C. Hernandez-Matas and X. Zabulis, "Super resolution for fundoscopy based on 3D image registration," in *EMBC*, 2014, pp. 6332–38.

[5] Y. Lin and G. Medioni, "Retinal image registration from 2D to 3D," *IEEE CVPR*, pp. 1–8, 2008.

[6] N. Meitav and E. Ribak, "Improving retinal image resolution with iterative weighted shift-and-add." *JOSA A*, vol. 28, no. 7, pp. 1395–402, 2011.

[7] G. Molodij, E. Ribak, M. Glanc, and G. Chenegros, "Enhancing retinal images by extracting structural information," *Optics Communications*, vol. 313, pp. 321–8, 2014.

[8] G. Troglio, A. Nappo, J. A. Benediktsson *et al.*, "Automatic Change Detection of Retinal Images," in *IFMBE*, 2009, pp. 281–4.

[9] J. P. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey." *IEEE Trans. on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[10] A. V. Cideciyan *et al.*, "Registration of High Resolution Images of the Retina," *SPIE Medical Imaging*, vol. 1652, pp. 310–22, Jun. 1992.

[11] G. Matsopoulos, N. Mouravliansky *et al.*, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. on Inf. Tech. in Biomed.*, vol. 3, no. 1, pp. 47–60, 1999.

[12] C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration." *IEEE trans. on Medical Imaging*, vol. 22, no. 11, pp. 1379–94, 2003.

[13] R. Poli, J. Kennedy, and T. Blackwell, "Particle Swarm Optimization," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *CVIU*, vol. 110, no. 3, pp. 346–59, 2008.

[15] O. Miksik and K. Mikolajczyk, "Evaluation of local detectors and descriptors for fast feature matching," in *ICPR*, 2012, pp. 2681–2684.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[17] P. Reel, L. Dooley, K. Wong, and A. Borner, "Multimodal retinal image registration using a fast principal component analysis hybrid-based similarity measure," *ICIP*, 2013.

[18] R. Navarro, J. Santamaría, and J. Bescós, "Accommodation-dependent model of the human eye with aspherics," *JOSA A*, vol. 2, no. 8, pp. 1273–81, 1985.

[19] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[20] S. Hinterstoisser *et al.*, "Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes," in *ACCV*, 2012, pp. 548–62.

[21] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–95, 1981.

---

[1]The results at http://iris.usc.edu/people/yupingli/research.html were utilized for the comparison.

[2]The implementation at http://www.vision.cs.rpi.edu/gdbicp/ was utilized.