An Experimental Evaluation of the Accuracy of Keypoints-based Retinal Image Registration

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

Abstract— **This work regards an investigation of the accuracy of a state-of-the-art, keypoint-based retinal image registration approach, as to the type of keypoint features used to guide the registration process. The employed registration approach is a local method that incorporates the notion of a 3D retinal surface imaged from different viewpoints and has been shown, experimentally, to be more accurate than competing approaches. The correspondences obtained between SIFT, SURF, Harris-PIIFD and vessel bifurcations are studied, either individually or in combinations. The combination of SIFT features with vessel bifurcations was found to perform better than other combinations or any individual feature type, alone. The registration approach is also comparatively evaluated against representative methods of the state-of-the-art in retinal image registration, using a benchmark dataset that covers a broad range of cases regarding the overlap of the acquired images and the anatomical characteristics of the imaged retinas.**

I. INTRODUCTION

Fundoscopy is a non-invasive diagnostic imaging procedure that allows for the acquisition of retinal images. One of its applications is the diagnosis and treatment of diseases related to microvascular circulation in the retina [1], such as hypertension and diabetes [2]. The analysis of retinal images is facilitated by the ability to register a test image with a reference one. Registration warps images to a common reference frame, so that a physical point on the retina is imaged at the same coordinates in all the warped images. Registration of retinal images can be used to create images of higher resolution and definition [3], [4], [5], to stitch images into a panoramic image of the retina, or a "mosaic" [6], [7], [8], and to facilitate longitudinal studies on retinopathy [9], [10].

Registration methods are based on the extraction of common information between the test and the reference images. Registration approaches can be categorized as global or local methods [11]. Global methods compare intensity patterns in images via correlation metrics. In retinal image registration, the few existing global methods are based on mutual information [12]. Local methods rely on localized features such as keypoints [13], [14], [15], [5], [16], [17], [7], [18], [19], [20] or vessel trees [21]. Local methods are more robust to local changes due to anatomical differences or illumination artifacts. As such, they are more popular compared to global methods. The local, keypoint-based approaches consitute the focus of this work.

[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

The type of keypoints used affects the accuracy of the registration. In this work, we present an investigation of this impact for the most widely-utilized features in the literature of retinal image registration. This investigation regards the registration framework presented in [17], [22]. This framework is chosen for its higher accuracy compared to state-of-the-art methods and for being representative of local methods, as it generically treats eye pose and shape similarly to the most successful state-of-the-art-methods [23], [24]. In addition, the framework instantiated to use the best performing keypoint combination is compared to characteristic, state-of-the-art local approaches.

The employed registration framework detects keypoints in both test and reference images and then matches them. By using the established correspondences (which may include some spurious ones) an initial pose estimation between the two views is obtained. This is achieved robustly using the RANSAC [25] algorithm and a spherical eye model. Using this initial estimate, the relative camera pose and the eye shape are refined simultaneously, to improve image registration. The multidimensional pose and shape space is efficiently searched by Particle Swarm Optimization [26], a stochastic method that iteratively evaluates hypotheses in the solution space.

II. KEYPOINT DETECTION AND EXTRACTION

In this work, we evaluate the impact of SIFT, SURF, Harris-PIIFD and Bifurcations as keypoints used by the selected retinal image registration framework. These keypoints have been selected for being the most widely employed for retinal image registration.

Scale-Invariant Feature Transform (SIFT) [27] is the milestone method in extracting characteristic points, or keypoints, in images. It is a key pillar in applications such as image mosaicing, robot mapping and navigation, object and gesture recognition and video tracking. SIFT are general purpose keypoints that are partially invariant to image translation, scaling, rotation as well as to illumination and affine distortion. Retinal registration methods utilizing SIFT keypoints are found in [13], [17].

Speeded Up Robust Features (SURF) [28] is another widely used method to detect and describe keypoints. It is partially invariant to image translation, scaling, rotation and illumination. Retinal registration methods employing SURF keypoints are found in [5], [16], [29].

Partial Intensity Invariant Feature Descriptor on Harris Corners (Harris-PIIFD) [30] is a method introduced with the

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

main purpose of finding keypoints in multimodal retinal image pairs. PIIFD is invariant to image rotation, and partially invariant to image intensity and perspective change. Retinal registration methods employing PIIFD are found in [30], [31], [29].

Bifurcations, or Y-features, are commonly used in retinal image registration [18], [7], [20], [19]. There exist different approaches in the literature regarding extraction and matching of bifurcations. In this work, bifurcations are extracted based on [19]. For each image from the pair, the retinal vessel tree is segmented by thresholding the response of Frangi's multiscale vessel enhancement filter [32]. The resulting binary image is skeletonized using Zhang's thinning method [33]. Bifurcations are then calculated by selecting vessel pixels that in their 8-neighborhood, have 3 or 4 non-adjacent vessel pixels. As bifurcations have no associated descriptors, in this work SIFT descriptors on the bifurcations are calculated. The resultant keypoints are matched conventionally as if they were SIFT keypoints.

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.

III. EXPERIMENTS AND RESULTS

One goal of the experiments is to investigate the retinal image registration accuracy of the framework in [17], [22], with respect to the type of utilized keypoints. A second goal is to compare the most accurate version of that framework to the state-of-the-art.

A. Datasets

We employed the publicly available[1] Fundus Image Registration (FIRE) dataset [34]. This dataset consists of 129 retinal images forming 134 image pairs, and classified into 3 different subcategories according to different types of applications.

Category $\mathcal{S}$ contains 71 image pairs with high overlap and no anatomical changes. Of the three categories, it presents the simplest image pairs to register. Large overlap leads to the possibility of extracting common information in a large area of the images, thus having more potential information to be utilized for the registration. Lack of anatomical changes facilitates the matching of keypoints, as the information describing them will be similar in both images. Indicative application for these image pairs is super resolution.

Category $\mathcal{P}$ contains 49 image pairs with small overlap and no anatomical changes. It presents increased registration difficulty compared to $\mathcal{S}$. As the overlap between the images is small, there is less area from which common information can be extracted, and less potential information to be used for the registration. Given the relative lack of texture in retinal images, that is a hurdle for keypoint extraction. Indicative application for these image pairs is mosaicing.

Category $\mathcal{A}$ contains 14 image pairs with high overlap and large anatomical changes. Similarly to category $\mathcal{P}$, $\mathcal{A}$ also presents an increased difficulty with respect to $\mathcal{S}$. In this case, the overlap is large, but there exist anatomical differences in the image pairs. These differences may appear in the form of increased vessel tortuosity, microaneurysms, cotton-wool spots, etc. As a result, features that are visible in one image, may be occluded in the other. If the same features are visible in both, they may look different, leading to the information describing them being different enough for the keypoints not to be matched. This may cause a low amount of corresponding keypoints, either in a part, or in the whole image. Indicative applications for these image pairs are longitudinal studies on retinopathy.

The top row of Figure 1 shows one image pair for each of the three categories. Ground truth is provided with the dataset in the form of corresponding points. Registration error in Figures 2 and 3 is indicated with a 2D plot inspired by relevant efforts in the field of evaluating object tracking methods [35]. The x axis of the plot corresponds to the value of an error threshold, and the y axis to the percentage of registered pairs with an average registration error lower than the corresponding threshold. This creates a continuous, monotonic curve, which shows success rate as a function of target accuracy. Thus, the selection of an arbitrary threshold is avoided. Error is also presented as the Area Under Curve (AUC) for the plot.

B. Experimental setup

Experiments were run in an Intel Core i7-4770 CPU @ 3.40GHz with 16GB of RAM memory and a NVIDIA GeForce GT 730 on Windows 7 Professional. The retinal image registration framework described in [17], [22] was utilized. The framework was implemented in C++ with openCV and CUDA tools. The framework was configured to use 3 independent PSO swarms, 300 generations per swarm and 10^4 particles per generation. In this configuration, the framework evaluates 9×10^6 hypotheses[2] for each image pair to be registered, and 1.81×10^{10} for the entire FIRE dataset for each of the 15 combinations of feature types considered in this study.

C. Keypoint comparison

This experiment aims to compare the impact of the keypoints described in Section II on the accuracy of the registration framework in [17], [22]. Table I shows the AUC for every feature combination when registering the FIRE dataset. Figure 2 shows the result plots for keypoints/keypoint combinations. For clarity, only three plots corresponding to the most accurate results are displayed.

For $\mathcal{S}$ the framework performs similarly, independently from the keypoints used. The exception to this is the utilization of Harris-PIIFD, as in this case the framework provides results one step below the rest. A weak performance in

[1] http://www.ics.forth.gr/cvrl/fire

[2] In this context, a hypothesis is a candidate solution to the problem, i.e., the parameters of the relative camera pose of the two retina views as well as the eye shape parameters. More details can be found in [17], [22].

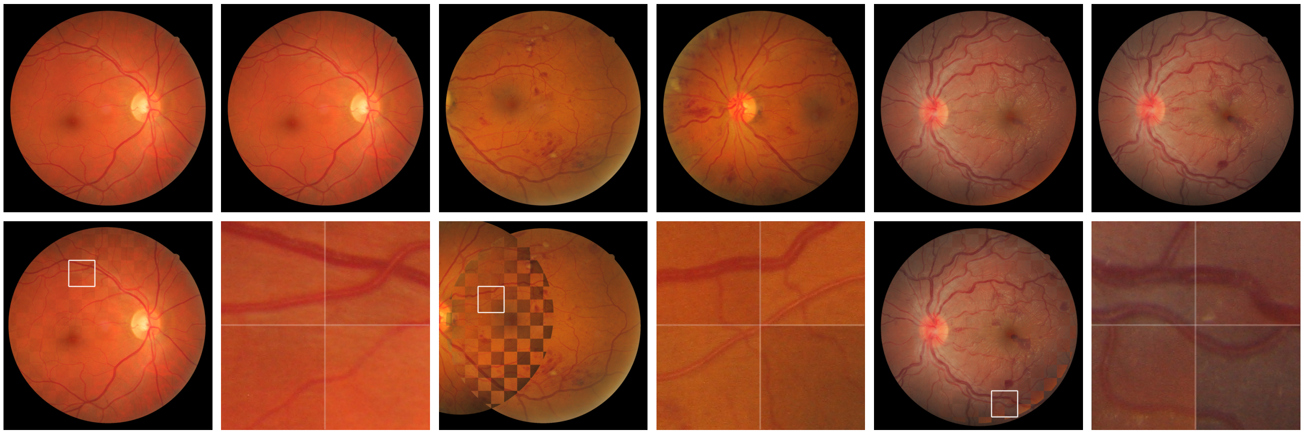

Fig. 1. Top row shows fundus image pairs from categories $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{A}$, respectively. Bottom row shows registration results. The collages show the reference and registered images, alternating in locations where they overlap. The marked region with solid white line indicates the magnified image detail shown on the right.

this category implies that Harris-PIIFD are relatively weaker keypoints. When employed, the framework exhibits a weaker performance for all categories of images in the dataset.

For $\mathcal{P}$, the framework provides the most accurate results when employing the combination of Bifurcations and SIFT keypoints. When studying the keypoints in isolation, the framework provides the best results using Bifurcations, followed by the utilization of SIFT.

For $\mathcal{A}$, the framework provides the best results when using the combination of Bifurcations and SIFT keypoints, closely followed by the combination of Bifurcations, SIFT and Harris-PIIFD. When utilizing the keypoints in isolation, the framework provides the most accurate results when employing SIFT, closely followed by Bifurcations.

In the FIRE dataset as a whole, and when studying keypoints in isolation, the framework provides the most accurate results when employing Bifurcations, followed by SIFT features. Overall, the best results are obtained when the framework utilizes the combination of these two types of keypoints.

In the initial instantiation of the utilized framework [17], [22] competitive results to the state-of-the-art were obtained using SIFT features. The results in this section, show that by utilizing this framework with other keypoints, i.e. SURF or Bifurcations, is also competitive to the state-of-the-art results. Moreover, it is found that the accuracy of the approach is increased, if keypoints that provide complementary information are combined. Bifurcations are clear points of interest located along vessels, which significantly contribute to the structure of the retinal image. Generic keypoint features respond to more generic image structure and are found in more locations, not only in retinal vessels. Out of the generic keypoint features, SIFT yielded the best results, while adding more (i.e. SURF) deteriorated performance. The experiments in this section confirm that combination of complementary features provides more registration evidence that leads to more accurate registration results, while retaining all other factors equal. Given that the utilized framework is representative of local methods, this behavior is expected to regard other local registration methods as well.

SIFT	SURF	PIIFD	Bifur.	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	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
×	×			0.953	0.423	0.526	0.712
×		×		0.951	0.396	0.503	0.699
×			×	0.958	**0.542**	**0.660**	**0.773**
	×	×		0.940	0.264	0.426	0.636
	×		×	0.956	0.404	0.489	0.703
		×	×	0.954	0.472	0.563	0.736
×	×	×		0.952	0.333	0.491	0.674
×	×		×	0.956	0.435	0.480	0.713
×		×	×	**0.959**	0.490	0.657	0.754
	×	×	×	0.954	0.400	0.474	0.699
×	×	×	×	0.956	0.409	0.514	0.707

TABLE I
AREA UNDER THE CURVE IN EACH FIRE CATEGORY FOR EVERY KEYPOINT COMBINATION. × INDICATES A KEYPOINTS TYPE USED FOR REGISTRATION. BOLD SHOWS THE HIGHEST SCORE.

D. *Comparison with state of the art*

In this experiment, the accuracy of the registration framework employing the combination of Bifurcations and SIFT is compared to three state-of-the-art methods. The first method is the previous iteration of the same framework [22], which utilized SIFT keypoints. The other two methods are GDB-ICP [23] and the original Harris-PIIFD framework [30]. These methods have been selected as both are widely applied in the retinal image registration field. GDB-ICP is used in [14], [18], [30], [36] and the original Harris-PIIFD framework in [31], [29]. For GDB-ICP, the C++ implementation provided by the authors[3] was used. For the original Harris-PIIFD framework, a MATLAB implementation was utilized.

Figure 3 and Table II show the result plot and the AUC, respectively. The framework utilizing the proposed keypoint

[3]`http://www.vision.cs.rpi.edu/gdbicp/exec/`

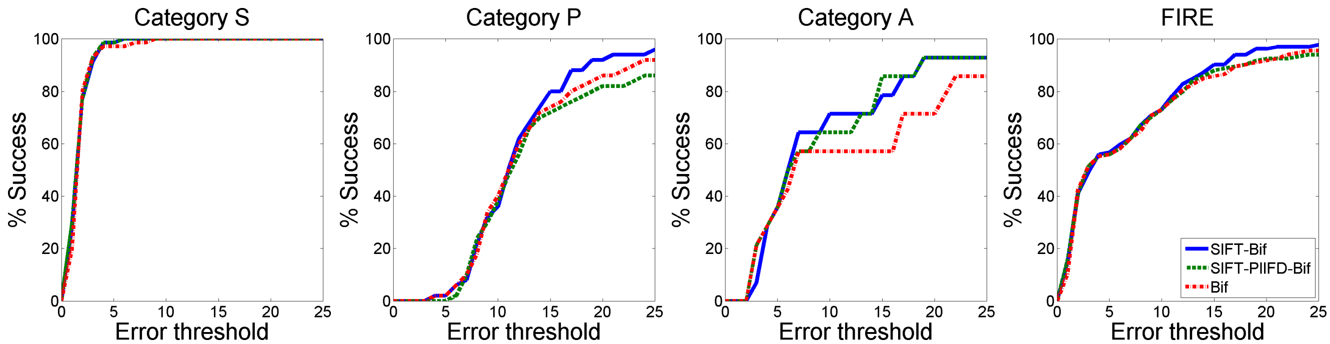

Fig. 2. Registration success for the following keypoint combinations: SIFT-Bifurcations, SIFT-PIIFD-Bifurcations and Bifurcations. The x axis marks the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.

Method	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	FIRE
H-M'17	**0.958**	**0.542**	**0.660**	**0.773**
H-M'16 [22]	0.945	0.443	0.577	0.721
Harris-PIIFD [30]	0.900	0.090	0.443	0.553
GDB-ICP [23]	0.814	0.303	0.303	0.576

TABLE II
AREA UNDER THE CURVE IN EACH FIRE CATEGORY FOR THE FOLLOWING REGISTRATION METHODS: PROPOSED (H-M'17), PREVIOUS ITERATION OF THE FRAMEWORK (H-M'16) [22], HARRIS-PIIFD [30] AND GDB-ICP [23]. BOLD SHOWS THE HIGHEST SCORE.

combination outperforms the original Harris-PIIFD framework [30] in every category, as well as the previous iteration of the framework in almost every condition. The proposed method outperforms GDB-ICP [23] except up to a certain threshold in $\mathcal{P}$.

IV. CONCLUSIONS

There exist several methods to extract corresponding keypoints in retinal images. The most popular ones are studied in this work, to analyze their impact in the accuracy of the retinal image registration framework proposed in [17], [22]. The publicly available[1] FIRE [34] dataset is used in the experiments. When studying keypoints in isolation, the framework provides the most accurate results when employing Bifurcations, followed by SIFT. Overall, the best registration results are obtained when the framework uses a combination of these two keypoints, showing that the combination of complementary keypoints allows for more accurate registration than when utilized in isolation. Additionally, the registration framework with the proposed keypoint combination is shown to perform more accurately than GDB-ICP and the original Harris-PIIFD framework, which are widely employed retinal image registration methods. To the best knowledge of the authors, retinal image registration methods that are based on keypoint correspondences employ a single type of keypoints. The registration framework used in the experiments is considered to be representative of local methods. Thus, the conclusion of combining certain types of keypoints to improve registration accuracy is expected to be useful towards improving other local registration methods, too.

ACKNOWLEDGMENTS

This research was made possible by a Marie Curie grant from the European Commission in the framework of the REVAMMAD ITN (Initial Training Research Network), Project 316990. It was also supported by the FORTH-ICS internal RTD Programme "Ambient Intelligence and Smart Environments". The authors would like to thank Dr. Z. Ghassabi for kindly providing the original Harris-PIIFD framework implementation.

REFERENCES

[1] M. Abramoff, M. Garvin, and M. Sonka, "Retinal imaging and image analysis." *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.

[2] 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.

[3] N. Meitav and E. Ribak, "Improving retinal image resolution with iterative weighted shift-and-add." *Journal of the Optical Society of America A*, vol. 28, no. 7, pp. 1395–402, 2011.

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

[5] C. Hernandez-Matas and X. Zabulis, "Super resolution for fundoscopy based on 3D image registration," in *IEEE EMBC*, Chicago, USA, 2014, pp. 6332–6338.

[6] 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 TPAMI*, vol. 24, no. 3, pp. 412–19, 2002.

[7] N. Ryan, C. Heneghan, and P. de Chazal, "Registration of digital retinal images using landmark correspondence by expectation maximization," *Elsevier Image and Vision Computing*, vol. 22, no. 11, pp. 883–98, 2004.

[8] P. Cattin, H. Bay, L. J. V. Gool, and G. Szkely, "Retina mosaicing using local features." in *MICCAI 2006. Proceedings, Part II*, vol. 4191, 2006, pp. 185–92.

[9] H. Narasimha-Iyer *et al.*, "Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences." *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 8, pp. 1436–45, Aug. 2007.

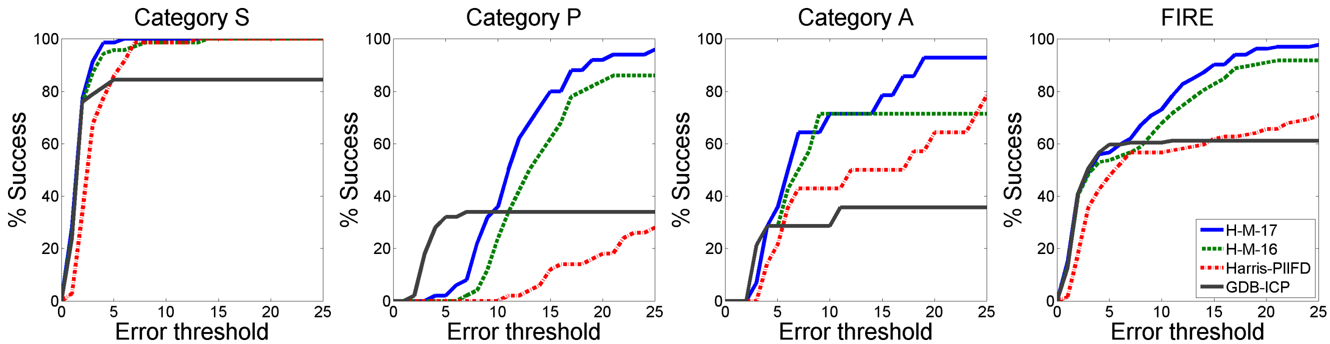

Fig. 3. Registration success for the following registration methods: Proposed (H-M'17), previous iteration of the framework (H-M'16) [22], Harris-PIIFD [30] and GDB-ICP [23]. Axes descriptions in Figure 2

[10] G. Troglio, M. Alberti, A. Benediksson, and G. Moser, "Unsupervised Change-Detection in Retinal Images by a Multiple-Classifier Approach," in *Multiple Classifier Systems*, 2010, pp. 94–103.

[11] L. G. Brown, "A survey of image registration techniques," *ACM Comp. Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[12] P. Legg, P. Rosin *et al.*, "Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation." *Elsevier Computerized Medical Imaging and Graphics*, aug 2013.

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

[14] C. Tsai, C. Li *et al.*, "The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence." *IEEE TMI*, vol. 29, no. 3, pp. 636–49, 2010.

[15] Li Tang, M. K. Garvin *et al.*, "Robust Multiscale Stereo Matching from Fundus Images with Radiometric Differences," *IEEE TPAMI*, vol. 33, no. 11, pp. 2245–2258, nov 2011.

[16] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal Image Registration Based on Keypoint Correspondences, Spherical Eye Modeling and Camera Pose Estimation," in *IEEE EMBC*, Milan, Italy, 2015, pp. 5650–5654.

[17] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. A. Argyros, "Retinal image registration under the assumption of a spherical eye," *Elsevier Computerized Medical Imaging and Graphics*, vol. 55, pp. 95 – 105, 2017.

[18] C. Stewart *et al.*, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration." *IEEE TMI*, vol. 22, no. 11, pp. 1379–1394, 2003.

[19] A. Chaudhry and J. Klein, "Ophthalmologic Image Registration based on shape-context: Application to Fundus Autofluorescence (FAF) images," *IASTED ICVIIP*, pp. 1–7, 2008.

[20] G. Matsopoulos, P. Asvestas, N. Mouravliansky, and K. Delibasis, "Multimodal Registration of Retinal Images Using Self Organizing Maps," *IEEE Transactions on Medical Imaging*, vol. 23, no. 12, pp. 1557–1563, dec 2004.

[21] G. Matsopoulos, N. Mouravliansky *et al.*, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, no. 1, pp. 47–60, 1999.

[22] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration through simultaneous camera pose and eye shape estimation," in *IEEE EMBC*, Aug. 2016, pp. 3247–3251.

[23] G. Yang, C. Stewart, M. Sofka, and C. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE TPAMI*, vol. 29, no. 11, pp. 1973–1989, 2007.

[24] K. M. Adal, P. G. van Etten *et al.*, "Accuracy Assessment of Intra- and Intervisit Fundus Image Registration for Diabetic Retinopathy Screening," *Investigative Ophthalmology and Visual Science*, vol. 56, no. 3, pp. 1805–1812, 2015.

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

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

[27] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Springer International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

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

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

[30] J. Chen, J. Tian, N. Lee *et al.*, "A partial intensity invariant feature descriptor for multimodal retinal image registration." *IEEE TBME*, vol. 57, no. 7, pp. 1707–18, Jul. 2010.

[31] Z. Ghassabi *et al.*, "An efficient approach for robust multimodal retinal image registration based on ur-sift features and piifd descriptors," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–16, 2013.

[32] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, *Multiscale vessel enhancement filtering*. MICCAI, 1998, pp. 130–137.

[33] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.

[34] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. A. Argyros, "FIRE: Fundus Image Registration Dataset," *Journal for Modeling in Ophthalmology*, 2017.

[35] B. Babenko, M. H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE TPAMI*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011.

[36] J. Zheng, J. Tian, K. Deng, X. Dai, X. Zhang, and M. Xu, "Salient feature region: a new method for retinal image registration." *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 221–32, 2011.