Retinal Image Registration through 3D Eye Modelling and Pose Estimation

Carlos Hernandez-Matas

June 2017

University of Crete
School of Sciences and Engineering
Computer Science Department

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Doctoral Thesis Committee: Antonis Argyros, Professor, University of Crete (Advisor)
Xenophon Zabulis, Principal Researcher, FORTH-ICS (Advisor)
Konstantinos Marias, Associate Professor, TEI of Crete
Georgios Tziritas, Professor, University of Crete
Manolis Lourakis, Principal Researcher, FORTH-ICS
Andrew Hunter, Professor, University of Lincoln
Alfredo Ruggeri, Associate Professor, University of Padova

The work reported in this thesis has been conducted at the Computational BioMedicine Laboratory (CBML) and the Computational Vision and Robotics Laboratory (CVRL) of the Institute of Computer Science (ICS) of the Foundation for Research and Technology–Hellas (FORTH). 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 number 316990. It was also supported by the FORTH-ICS internal RTD Programme "Ambient Intelligence and Smart Environments".

UNIVERSITY OF CRETE
DEPARTMENT OF COMPUTER SCIENCE

Retinal Image Registration through 3D Eye Modelling and Pose Estimation

Dissertation submitted by
Carlos Hernandez-Matas

In partial fulfilment of the requirements for the
Doctor of Philosophy Degree in Computer Science

Author: Carlos Hernandez-Matas

Committee Member: Antonis Argyros, Professor, University of Crete

Committee Member: Xenophon Zabulis, Principal Researcher, FORTH-ICS

Committee Member: Konstantinos Marias, Associate Professor, TEI of Crete

Committee Member: Georgios Tziritas, Professor, University of Crete

Committee Member: Manolis Lourakis, Principal Researcher, FORTH-ICS

Committee Member: Andrew Hunter, Professor, University of Lincoln

A. Ruggeri

Committee Member: Alfredo Ruggeri, Associate Professor, University of Padova

Department Chairman: Angelos Bilas, Professor, University of Crete

Heraklion, June 2017

Acknowledgements

The path to obtaining a doctorate is long and arduous, filled with hurdles to overcome. Fortunately, it is rare to be alone in this road, as it is usually the case that we have the support of our colleagues, friends, and of course, family. This page is for them.

I would like to thank Dr. Xenophon Zabulis, for giving me the chance to join the REVAMMAD project, and Prof. Antonis Argyros, for accepting me as a Ph.D. student and inviting me to join the Computational Vision and Robotics Laboratory (CVRL). They steered my research every time I needed guidance, and showed me more patience than I actually deserved. I am also very grateful to Assoc. Prof. Konstantinos Marias for welcoming me to the Computational BioMedicine Laboratory (CBML) and to Assoc. Prof. Alfredo Ruggeri for doing the same at the Biomedical Imaging Laboratory (BioImLab). Let us not forget the members of CVRL, CBML and BioImLab, as they provided me with great support during this time. And many thanks to the members of my thesis committee, in particular Dr. Manolis Lourakis, as his invaluable feedback helped improve the contents of this thesis. Additional thanks to Dr. Areti Triantafyllou and Dr. Panagiota Anyfanti, for their assistance with the clinical side of the project. Thank you.

I would also like to thank the friends I made during this period, as well as the friends I had to leave behind every time I have moved to a new city. Without their support, this would have been an impossible task. They are simply too many to name all of them. Thank you.

And of course my parents Pitu and Manolo, my brother Javi, my nieces Dani and Noe, and the rest of my loving family, who understand the sacrifices I had to make during the years in the pursuit of my dream. And I want to thank Ruth, who has a key role in my life, and who has believed in me and supported me at every step. Thank you.

Finally, if you are reading this, you are making the effort put into this work to be worth it. So thank you to you too.

Abstract

The in vivo assessment of small vessels can promote accurate diagnosis and progression monitoring of diseases related to vasculopathy, such as hypertension and diabetes. Given that the human eye retina contains small vessels that can be directly imaged via fundoscopy, the analysis of retinal structures becomes very important for non-invasive approaches. This is also important for the diagnosis of illnesses that affect eyesight, such as macular edema, age-related macular degeneration or glaucoma. This analysis can be greatly facilitated by accurate retinal image registration.

Image registration is applied upon a pair of images, the reference and the test image. The goal is to warp the test image so that it images retinal points at the same 2D locations as the reference image. This is a challenging task, mainly for two reasons. The first is related to the perspective distortions due to the curved shape of the retina and change of the camera pose relative to the eye between image acquisitions. The second relates to potential changes in the retina that occur due to retinopathy between temporally distant image acquisitions. In addition, the nature of the application demands for high registration accuracy.

Registration methods may benefit from knowledge of the type of images to be registered. In this work, we proposed a registration framework that simultaneously estimates the relative pose of the cameras that acquired the retinal images as well as the shape and the pose of the eye. The proposed framework, which has been made publicly available, is evaluated quantitatively and is shown to outperform state-of-the-art methods.

In the context of this work, we also developed a set of tools for generating realistic 3D eye models. These tools were used to render synthetic retinal image pairs, utilized for testing and evaluating the proposed registration approach. Additionally, FIRE, a dataset comprised of pairs of real retinal images has been compiled and made publicly available. FIRE consists of three types of images, each one covering different challenges in retinal image registration. To enable the experimental, quantitative evaluation of the accuracy of a registration method, FIRE is annotated with ground truth point correspondences.

In this work, we also explored the suitability of the proposed registration framework for applications such as longitudinal studies, image mosaicing and super resolution. Additionally, the fitness of the framework for performing eye shape estimation is studied. Pertinent experiments show encouraging results as well as ample room for further improvement.

Περίληψη

Η μελέτη των λεπτών αιμοφόρων αγγείων μπορεί να βοηθήσει την ακριβή διάγνωση και την παρακολούθηση νόσων που σχετίζονται με αγγειοπάθεια, όπως η υπέρταση και ο διαβήτης. Η ευκολία απεικόνισης τέτοιων αγγείων στον αμφιβληστροειδή χιτώνα μέσω βυθοσκόπισης καθιστά την ανάλυση των δομικών χαρακτηριστικών του ιδιαίτερα σημαντική. Η ανάλυση αυτή είναι επίσης σημαντική για τη διάγνωση ασθενειών που επηρεάζουν την όραση, όπως το οίδημα της ωχράς κηλίδας, η ηλικιακή εκφύλιση της ωχράς κηλίδας και το γλαύκωμα. Η ανάλυση των δομικών χαρακτηριστικών του αμφιβληστροειδή μπορεί να διευκολυνθεί σημαντικά από την ακριβή αντιστοίχιση (registration) των εικόνων του.

Η αντιστοίχιση εφαρμόζεται σε ένα ζεύγος εικόνων του αμφιβληστροειδή, την εικόνα αναφοράς και την εικόνα ελέγχου. Στόχος είναι ο γεωμετρικός μετασχηματισμός της εικόνας ελέγχου έτσι ώστε τα σημεία της να απεικονίζονται στις ίδιες 2Δ συντεταγμένες με την εικόνα αναφοράς. Η επίτευξη αυτού του στόχου συναντά δυσκολίες για δύο κυρίως λόγους. Ο πρώτος σχετίζεται με τις προοπτικές παραμορφώσεις που προκύπτουν από το καμπύλο σχήμα του αμφιβληστροειδή χιτώνα σε συνδυασμό με την αλλαγή της θέσης και του προσανατολισμού της κάμερας ως προς τον οφθαλμό (σχετική πόζα) κατά την πρόσληψη των εικόνων αναφοράς και δοκιμής. Ο δεύτερος σχετίζεται με πιθανές μεταβολές στη δομή του αμφιβληστροειδή μεταξύ χρονικά απομακρυσμένων προσλήψεων των εικόνων αναφοράς και ελέγχου, εξαιτίας κάποιας αμφιβληστροειδοπάθειας. Επιπρόσθετα, η φύση της εφαρμογής απαιτεί η αντιστοίχιση να γίνεται με μεγάλη ακρίβεια.

Οι μέθοδοι αντιστοίχισης μπορούν να επωφεληθούν από τη γνώση του είδους των εικόνων προς αντιστοίχιση. Σε αυτή την εργασία, προτείνεται ένα μεθοδολογικό πλαίσιο αντιστοίχισης εικόνων που ταυτόχρονα εκτιμά τη σχετική πόζα μεταξύ των δύο όψεων του αμφιβληστροειδούς καθώς και την πόζα και το σχήμα του οφθαλμού. Η ποσοτική αποτίμηση του προτεινόμενης μεθοδολογίας αποδεικνύει πειραματικά πως αυτή υπερτερεί σε ακρίβεια αντιστοίχισης έναντι των σύγχρονων μεθόδων.

Στο πλαίσιο αυτής της εργασίας αναπτύχθηκαν επίσης υπολογιστικά εργαλεία για τη σύνθεση ρεαλιστικών 3Δ μοντέλων του οφθαλμού που χρησιμοποιήθηκαν για τη σύνθεση εικόνων του αμφιβληστροειδούς χιτώνα από διαφορετικές όψεις, με τελικό στόχο τον έλεγχο και την αξιολόγηση της προτεινόμενης προσέγγισης. Επιπρόσθετα, συντάχθηκε και έγινε δημόσια διαθέσιμο το FIRE, ένα σύνολο δεδομένων που αποτελείται από ζεύγη πραγματικών εικόνων του αμφιβληστροειδούς. Το FIRE αποτελείται από τρείς τύπους ζευγών εικόνων, καθένας από τους οποίους παρουσιάζει διαφορετικού είδους δυσκολίες και προκλήσεις ως προς την αντιστοίχισή τους. Προκειμένου να γίνει εφικτή η πειραματική, ποσοτική αποτίμηση της ακρίβειας μιας μεθόδου αντιστοίχισης, κάθε ζεύγος εικόνων του FIRE επισημειώθηκε με πραγματικές αντιστοιχίες σημείων.

Στην εργασία αυτή διερευνήθηκε επίσης η καταλληλότητα του προτεινόμενου πλαισίου αντιστοίχισης για να υποστηρίξει εφαρμογές όπως οι διαχρονικές συγκριτικές μελέτες εικόνων αμφιβληστροειδούς, η δημιουργία μωσαϊκών από εικόνες και η δημιουργία εικόνων υψηλής χωρικής ανάλυσης μέσω υπερδειγματοληψίας. Επιπρόσθετα, μελετήθηκε η καταλληλότητα του πλαισίου αυτού για την εκτίμηση των παραμέτρων του σχήματος του οφθαλμού. Τα σχετικά πειράματα δείχνουν ενθαρρυντικά αποτελέσματα καθώς και την ύπαρξη σημαντικών περιθωρίων για περαιτέρω βελτίωση.

Contents

List of Tables

List of Figures

Abbreviations

AUC Area Under Curve. 56, 60, 62–67

FIRE Fundus Image Registration. 6, 47–54, 56, 57, 59, 60, 62–67, 71–73, 83, 87, 88

FOV Field of View. 5, 13, 18, 19, 24, 44, 48, 51, 54, 57, 73, 78, 85

GDB-ICP Generalized Dual Bootstrap-Iterative Closest Point. 23, 67–69, 73, 77–80, 95

MSE Mean Square Error. 73, 78, 80

OCT Optical Coherence Tomography. 3, 14

PIIFD Partial Intensity Invariant Feature Descriptor. 19, 31, 32, 41, 60, 62, 67–69, 73, 77–80, 96–98

PnP Perspective-n-Point. 33, 34, 41

PSNR Peak Signal-to-Noise Ratio. 73, 77, 79

PSO Particle Swarm Optimization. 34, 35, 39, 43–46, 59, 62–65, 68, 69, 86

RANSAC Random Sample Consensus. 33, 34, 39, 41, 43, 44, 46, 59, 61–63, 65, 68

REMPE Registration through Eye Modelling and Pose Estimation. 39, 59, 68, 71, 79, 80, 87, 88

REVAMMAD Retinal Vascular Modeling, Measurement And Diagnosis. 88, 89

RIR Retinal Image Registration. 3–6, 12, 17, 19, 23–25, 31–33, 46, 47, 49, 50, 56, 57, 59, 60, 68, 71, 73, 77, 79, 86, 87, 93

SIFT Scale-Invariant Feature Transform. 19, 26, 31, 32, 39, 41, 46, 54, 59–62, 69, 95, 97

SLO Scanning Laser Ophthalmoscope. 14, 54

SNR Signal-to-Noise Ratio. 73, 77, 79

SR Super Resolution. 5, 6, 24, 25, 45, 50, 52, 54, 71, 73, 77–80

SSIM Structural Similarity. 73, 77, 80

SURF Speeded Up Robust Features. 19, 31, 32, 39, 54, 61, 62, 97–99

Nomenclature

F_0 Reference image for the registration. 17, 39, 42, 45, 68, 79

F_t Test image for the registration. 17, 39, 42, 45, 68, 79

K Intrinsic camera matrix. 37, 39, 49

$\mathbf{A}$ Diagonal matrix representing the semi-axes of the eye model. 36, 37, 42, 45, 49, 61, 69, 78–80

$\mathbf{Q}$ Rotation matrix for the eye model semi-axes. 36, 37, 42, 45, 49, 61, 69

$\mathbf{R}$ Test camera relative rotation. 36, 37, 41–43, 45, 49, 56, 78, 79

$\mathbf{S}$ Parameterization of eye shape and camera pose. 36, 37, 41, 42, 45, 46, 49, 55, 56, 59, 63, 77–79

$\mathbf{t}$ Test camera relative translation. 36, 37, 41–43, 45, 49, 56, 78, 79

$\mathcal{A}$ Category $\mathcal{A}$ of the FIRE dataset. 50–53, 60–64, 66, 67, 71, 72, 83, 84

$\mathcal{E}$ Retinal surface of the eye model. 36, 37, 39, 42, 45, 49, 68, 69

$\mathcal{P}$ Category $\mathcal{P}$ of the FIRE dataset. 50–53, 60, 62–64, 66, 67, 71, 73, 83, 84

$\mathcal{S}$ Category $\mathcal{S}$ of the FIRE dataset. 50–54, 60, 62–64, 66, 67, 71, 72, 83, 84

ρ Eye radius for the spherical model. 35, 37

a First semi-axis of the eye model. 36, 39, 42, 44

b Second semi-axis of the eye model. 36, 39, 42, 44

c Third semi-axis of the eye model. 36, 39, 42, 44

fov Field of view of the camera. 35, 37

f Focal distance of the camera. 35, 37

g Number of generations for a PSO process. 34, 35, 59, 63, 64

ltc Camera lens to cornea distance. 35, 37

l Amount of images used for super resolution. 25, 77

n Super resolution scaling factor. 25, 77

p Number of particles per generation for a PSO process. 34, 35, 59, 63, 64

p Focal distance of the camera. 35, 37

r_ω Rotation of the camera on the z axis. 36, 42

r_ϕ Rotation of the camera on the y axis. 36, 42

r_θ Rotation of the camera on the x axis. 36, 42

r_a Rotation of semi-axis a of the eye model. 36, 39, 42, 44

r_b Rotation of semi-axis b of the eye model. 36, 39, 42, 44

r_c Rotation of semi-axis c of the eye model. 36, 39, 42, 44

r Image radius in pixels. 35, 37

s Number of independent iterations for the registration. 44, 59, 66

t_x Relative camera translation on the x axis. 36, 42

t_y Relative camera translation on the y axis. 36, 42

t_z Relative camera translation on the z axis. 36, 42

$\mathcal{SR}1$ Dataset 1 for super resolution. 50, 54, 55, 77, 78

$\mathcal{SR}2$ Dataset 2 for super resolution. 50, 54, 55, 77–80

Part I

Introduction

Chapter 1

Introduction

Assessment of small vessels in vivo can promote accurate diagnosis and progression monitoring of diseases with strong vasculopathy, such as hypertension [1] and diabetes [2]. The analysis of retinal structures is also important for the diagnosis of illnesses that affect the eyesight, such as macular edema, age-related macular degeneration or glaucoma [3]. Of all the organs within the human body, the eye, particularly the retina, provides an easily accessible way to non-invasively estimate the microvascular status via fundoscopy [3]. Retinal images can be acquired with several imaging devices, such as fundus camera, or Optical Coherence Tomography (OCT) [3].

The analysis of fundus images can be greatly facilitated by Retinal Image Registration (RIR). Image registration is applied upon a pair of images, the reference and the test one. The goal is the spatial warping of the target image so that its points are imaged at the 2D coordinates of the corresponding points in the reference image. The reference and test images may differ with respect to the viewpoint, the time and the image acquisition device. RIR is the application of image registration to retinal fundus images.

1.1 Medical motivation

There are several medical applications of RIR. Images acquired during the same examination are not expected to have significant anatomic changes. If the image pair presents significant overlap, images can be combined to generate images of higher resolution and definition [4]–[6], enabling more accurate measurements of the vessel structure such as Arteriolar-to-Venular Ratio [7]. In contrast, images with minor overlap can be combined into mosaics that image larger retinal areas [8]–[10]. The smaller the overlap, the lower the number of images needed to image a large area of the retina, thus increasing examination efficiency. Additionally, image pairs can be utilized for reconstructing the surface of the retina [11]–[14].

Images acquired during different moments in time can be registered and employed for performing longitudinal studies of the retina [15], [16]. This allows to monitor health status and disease progression of the patient over different time-points. It may represent an alternative method of 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.

Hypertension and hypertensive retinopathy are example medical conditions where assistive tools for the early diagnosis are of great clinical importance. Hypertension [17], also known as high blood pressure or raised blood pressure, can lead to heart attack, heart failure, strokes, aneurysms, kidney failure, blindness and cognitive impairment. Vessels can also develop aneurysms, clog and burst. As estimated by the World Health Organization [17], in 2013, there were about 1 billion people with hypertension in the world. It causes approximately 9.4 million deaths every year. That accounts for at least 45% of deaths due to heart disease. Hypertension is more present in countries with low and middle incomes, and most of the people with the illness actually have no visible symptoms. If the sickness is diagnosed early, the risk of heart attack, heart failure, stroke and kidney failure can be minimized.

Diabetes and diabetic retinopathy are other example medical conditions that benefit from these assistive analysis tools. Complications due to diabetes [18] include heart attack, stroke, kidney failure, leg amputation, vision loss and nerve damage. In 2014, the World Health Organization estimated [18] that 422 million adults were living with diabetes, and that it caused 1.5 million deaths in 2012.

Additional illnesses that benefit from retinal analysis are retinopathy of prematurity, age-related macular edema, photic retinopathy and radiation retinopathy [19].

1.2 Technical challenges

Despite the significant amount of work in the RIR field, the problem remains open and presents several technical challenges that must be addressed:

- *Different viewpoints*: Images acquired may image drastically different areas of the retina. This means that the overlapping area of an image pair might be small, thus reducing the amount of potential information than can be utilized by a RIR algorithm to achieve accurate registration.

- *Projection distortion*: The surface of the retina is of curved shape, as the shape of the eye can be approximated by an ellipsoid. Thus, projecting the retinal fundus onto a flat image introduces radial distortion.

- *Changes due to retinopathy*: Images may be acquired at different time instances. As such there may exist differences in the retina due to the appearance, progression or remission of retinopathy. Thus, the information located in the same eye area in two different images may vary. A microaneurysm may appear, vessels may get thinner or display increased tortuosity, or cotton-wool spots may cover previously visible landmarks.

- *High accuracy*: The purpose of RIR is to support image analysis by doctors and clinicians. As such, high accuracy on the registration is required to facilitate relevant analyses. High accuracy is also required for supporting applications based on registration such as mosaicing or Super Resolution (SR).

- *Acquisition artifacts*: Acquiring retinal images is a complicated task. Images are acquired through the pupil, which is a small aperture at the front of the eye. This leads to artifacts produced during the image acquisition. The iris and eyelashes may partially block the lighting source, producing illumination variance. Additionally, they may create artifacts in the final acquired image.

- *Different modalities*: Retinal images may be acquired in a range of modalities. Different modalities may display image information in different ways, and they might capture complementary information. As an example, angiograms are grayscale images in which microaneurysms are depicted more prominently than in color fundus images.

- *Different acquisition devices*: Retinal images within the same modality may be acquired with a range of devices. These devices may present different Field of View (FOV), resolution and contrast. Additionally, typically these devices are not calibrated.

1.3 Thesis contributions

In this work, several contributions on the topic of RIR are presented:

- A publicly available registration framework (`http://www.ics.forth.gr/cvrl/rempe`) based on simultaneous camera pose and eye shape estimation under several eye models. The performance of the framework is evaluated and compared to state-of-the-art registration methods, both general purpose and domain specific for retinal images. Additionally, the suitability of the framework for RIR applications is explored.

- A set of tools for generating synthetic, realistic 3D eye models, developed for testing and evaluation of the proposed approach. These tools also allow to create 2D retinal images from different views and eye models.

- A publicly available dataset (`http://www.ics.forth.gr/cvrl/fire`) comprised of pairs of real retinal images, as well as a ground truth and an evaluation protocol. This dataset consists of three types of images covering challenges in retinal image registration. This facilitates the comparison of RIR frameworks within the community.

1.4 Thesis outline

In Chapter 2 'The human eye', the structure of the eye, and in particular of the retina, is presented. Symptoms of retinopathy, which can cause physiological changes in the retina, are described. Additionally, a range of devices to acquire retinal images is presented.

In Chapter 3 'Retinal image registration', the problem of image registration is defined. Classifications for registration methods according to diverse criteria are presented. Related work in the field of RIR is presented and classified.

In Chapter 4 'Applications', an overview of applications that stem from RIR is provided. Additionally, for each of the applications related work is presented.

In Chapter 5 'Features, computational tools and models', the building blocks of this work are introduced. It covers keypoint extraction and matching, optimization methods, eye models and the geometry utilized.

Chapter 6 'REMPE registration framework' is the core part of the thesis. A RIR framework based on simultaneous camera pose and eye shape estimation is described.

In Chapter 7 'Datasets', publicly available retinal image datasets are analyzed. A framework to generate 3D and 2D synthetic retinal images from an existing retinal texture is presented. Additionally, the Fundus Image Registration (FIRE) dataset, a dataset publicly released with the aim to evaluate RIR is introduced, as well as a dataset for evaluating SR.

Chapter 8 'REMPE configuration experiments', presents experiments that support the selection of the framework's parameters as well as the evaluation of its performance. The method is also compared to state-of-the-art RIR methods.

In Chapter 9 'REMPE registration applications', the proposed RIR method is utilized as a stepping stone in applications based on image registration. Its utilization in applications such as longitudinal studies, mosaicing, SR and eye shape estimation is analyzed.

In Chapter 10 'Future work', possible research directions that could benefit the registration framework are discussed.

Chapter 11 'Discussion', summarizes the contributions of this work and presents its impact.

Finally, Appendix A 'Registration methods', presents an overview of existing registration methods that goes in more detail than the one presented in Chapter 3.

Chapter 2

The human eye

The eye is one of the most important sensory organs for humans, as it is the one we rely on the most for interacting with our environment. It works by detecting light and converting it into information through which we can perceive our surroundings, in short and long distances. Its general shape is that of an ellipsoid with axes of similar length, approximating a sphere.

2.1 Eye structure

The eye is a complex structure composed by 8 main elements, as shown in Figure 2.1.

- The cornea is a transparent layer at the front of the eye that covers the iris and the anterior chamber.
- The iris is a circular structure of the eye that controls the amount of light that reaches the retina.
- The crystalline lens is a biconvex structure in the eye that refracts the light that reaches the retina.
- The anterior and posterior chambers comprise the space located between the crystalline lens and the cornea. They are filled with aqueous humor, a fluid similar to plasma.
- The vitreous humor is located between the lens and the retina and consists of a gelatinous transparent mass.
- The optic disc is the entry and exit point of the eye for the blood vessels and nerves.
- The retina is a light-sensitive layer of tissue.

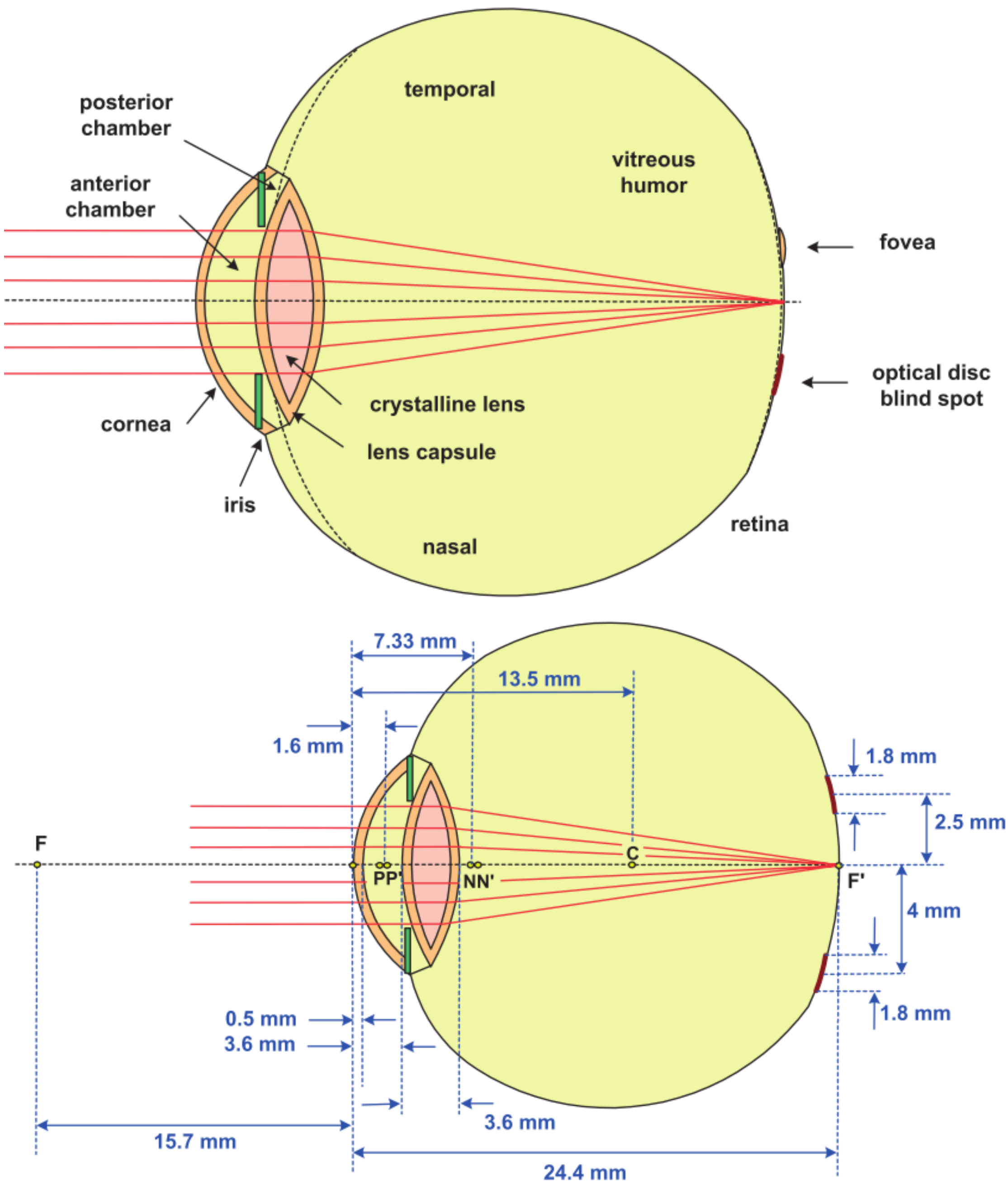


Figure 2.1: Diagrams of the eye structure [20].

Figure 2.2: The retina.

The retina is the main element of focus in this work, as our aim is to register retinal images. However, the rest of the elements of the eye play an important role during acquisition of retinal images. The cornea, crystalline lens, anterior chamber and vitreous humor create an intricate system of lenses that has to be crossed by the light captured by the image acquisition devices. Additionally, the iris controls the aperture of that system of lenses. As such, a realistic eye model must take these elements into consideration.

2.2 The retina

The retina is a layer of photosensitive tissue that coats the inner side of the eye. It consists of a structure composed by several layers of neurons, which are connected by synapses. Photoreceptor cells are the only neurons that are sensitive to light, and they can be classified into three types: rods, cones and ganglion cells. The network of rods allows to observe low resolution perception in black and white. Cones, on the other hand, allow for high resolution color perception. Finally, ganglion cells are important for reflexive responses to bright light. A frontal view of the retinal fundus is shown in Figure 2.2.

Blood reaches the retina in two different ways, both supplied by the ophthalmic

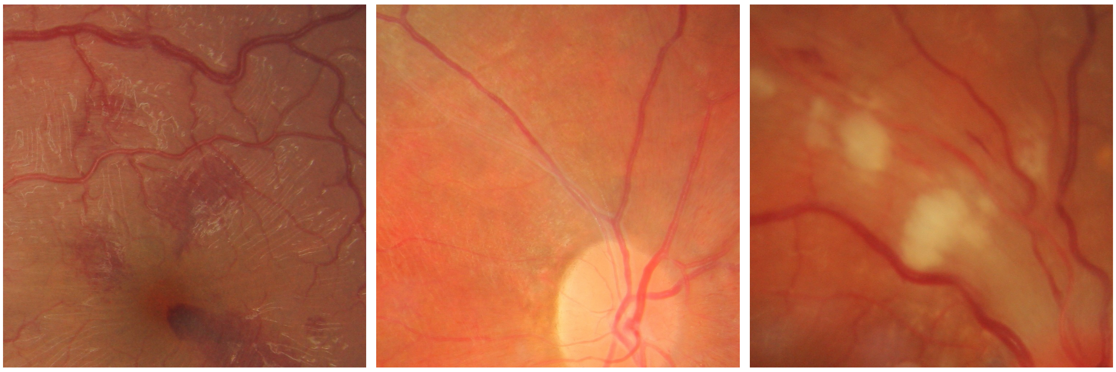


Figure 2.3: Retinopathy: Increased tortuosity and hemorrhages (left); arteriolosclerosis (middle); and cotton-wool spots, increased vascular tortuosity and hemorrhages (right).

artery. The central artery and vein of the retina enter the eye through the optic disc, and bifurcate several times into arterioles and venules that distribute the blood to the inner layers of the retina. The middle and outer layers of the retina are supplied blood via the uveal circulation, in which a series of branches of the ophthalmic artery penetrate the globe without doing so through the optic disc. The vessel tree present in the inner layer in the retina provides most of the potential information that can be exploited by registration methods.

2.2.1 Retinopathy

Retinopathy consists of the presence of damage in the retina. It can have several causes, and manifests in different forms. While retinopathy is of high importance to clinicians in the study of diseases, it is also quite relevant for Retinal Image Registration (RIR). In images acquired at different instances of time, there may exist differences in the retina due to retinopathy, increasing the difficulty for registering such images. A non-comprehensive list of retinopathy symptoms are described in this section, with some shown in Figure 2.3. A single symptom can be caused by several diseases [1], [2].

- Arteriolosclerosis consists of generalized or focalized arteriolar narrowing. It is associated with vessel wall thickening.
- Arteriolar wall opacity, or commonly known as copper (central light reflex takes most of the width of the vessel) or silver wiring (it takes all the width). It is the result of arteriolosclerosis.
- Arteriovenous nicking happens when an artery is crossing a vein. As the space they have is limited, the vein gets compressed.
- A microaneurysm consists of the swelling in the side of a blood vessel. It is usually a precedent to hemorrhages.

- A hemorrhage is an escape of blood from a ruptured blood vessel. It is usually blot, dot, or flame shaped.

- Soft exudates, or cotton-wool spots, are swollen nerve fibers in the layer surface of the retina. They look like blurry white-yellowish patches in the retina.

- Hard exudates are similar to cotton-wool spots, but more prominent and well defined. They come from leakage from pre-capillary arterioles

- Optic disc swelling consists of the inflammation of the optic disc.

- Increased vascular tortuosity is a vascular anomaly that consists of the elongation of blood vessels in such a way that the curves of the vessel become more prominent and tortuous.

2.3 Imaging techniques

There are several methods and imaging devices to acquire retinal images [3]. Different image acquisition devices may image different characteristics of the retina, and the acquired images may have diverse properties with respect to resolution, Field of View (FOV), or chromatic scale. Due to this, the acquisition device utilized is relevant to clinicians and medical experts, as a particular device may be required for a particular type of analysis, such as angiographies to study blood flow [3]. Furthermore, the combination of images acquired with different devices may be needed. Examples of images acquired with these devices are shown in Figure 2.4.

2.3.1 Fundus images

Fundus images are 2-dimensional representations of the 3-dimensional tissues of the retina. They can be obtained with a range of types of acquisition devices:

- In color fundus images [21] obtained with a fundus camera, the retina is illuminated with white light, obtaining a full color image. The intensity of the pixels show the amount of reflected light in the Red, Green and Blue wavebands.

- In red-free fundus images [22], the light is filtered to remove the red color. This improves the contrast on the images, allowing to better differentiate vessels and other structures. The most frequent way to obtain a red-free image is to use only the green channel, the blue channel or a combination of both from a color fundus image [6], [23]–[32]. This alternative is frequently used in the preprocessing of retinal images, as these type of images provide higher contrast than color images.

- Hyperspectral imaging [33] captures spectral information in wavelengths that range from ultraviolet to infrared. It has the advantage of depicting tissues that are not visible under white light.

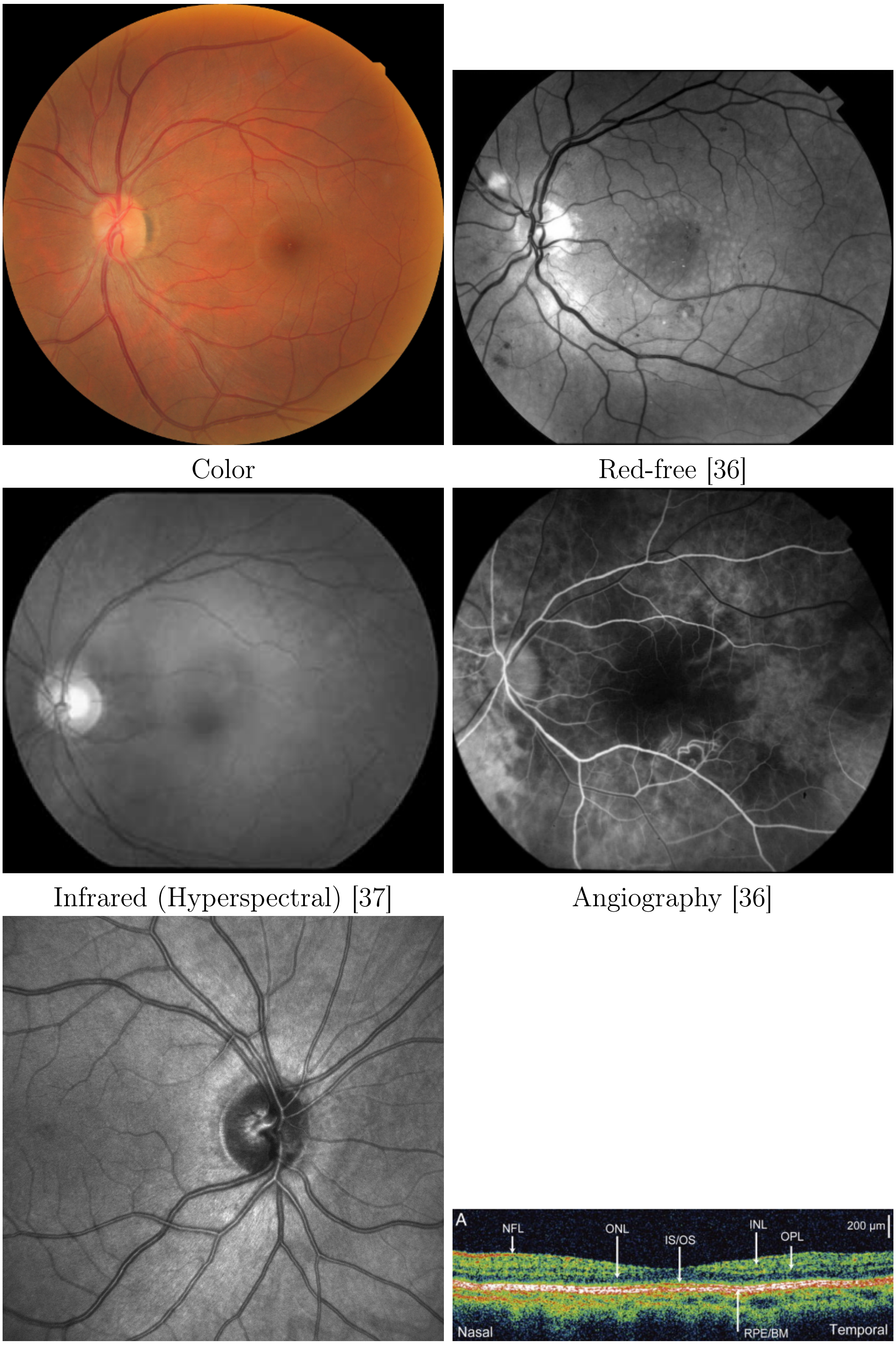
- In an angiography [34], an intravenous injection with a fluorescent dye is performed, typically in the arm of the subject. This fluorescent dye allows the vessels to be clearly seen under a specific light, which depends on the dye used. Typical dyes used are fluorescein and indocyanine. The intensity of the pixels shows the amount of emitted photons by the dye injected in the subject. Due to the nature of these methods, in which the dye circulates through the vessels, typically a sequence if images is taken, which also helps study the blood flow.

- Scanning Laser Ophthalmoscope (SLO) [35] uses confocal laser scanning microscopy to image the retina. In this technique, coherent light from the laser passes through a light source pinhole aperture that is perpendicular both to the point to scan and another pinhole aperture, which is in front of the detector. The laser light is reflected by a dichromatic mirror and scanned in a defined focal plane in the specimen. Fluorescence emitted by the specimen on the same focal plane passes through the mirror and reach the photodetector through the pinhole aperture.

2.3.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) is a high resolution 3D imaging system. It is based on low coherence interferometry. There are three main OCT principles for performing A-scan measurements. They are Time-domain OCT, Time encoded frequency domain OCT and Spectral-domain OCT.

While the work presented here focuses on frontal views of the retina, OCT devices capture cross sections of the retina, as shown in Figure 2.4. Thus this type of imaging is out of the scope of this research.



Color | Red-free [36] | Infrared (Hyperspectral) [37] | Angiography [36] | SLO | OCT B-scan [38]

Figure 2.4: Retinal imaging samples.

Chapter 3

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 are located at the same 2D locations in the reference frame of F_0. As described in Section 1.2, Retinal Image Registration (RIR) proves to be challenging due to image pairs having the potential of being from different modalities, acquired with different devices, showing changes due to retinopathy, having different viewpoints and presenting acquisition artifacts and projection distortion.

A wide range of registration methods [39] exists in the literature. Domains such as medical imaging [40] and remote sensing [41] have specific registration methods tailored to them. This is due to the fact that knowledge of the types of image pairs and anatomical features to be registered can be exploited to obtain better registration performance. In the particular case of RIR, the information provided by the vessel trees or the fact that the retinal surface can be approximated to a part of an ellipsoidal surface has been utilized.

3.1 Classification of registration algorithms

Registration algorithms can be classified according to a variety of criteria such as the the type of information used, the transformation model employed or the image modality.

3.1.1 Spatial vs frequency domain

Image registration is performed by exploiting information in the common regions of the two images. This information can be retrieved either in the frequency [42] or in the spatial domains [6], [9], [11], [12], [23]–[32], [36], [37], [43]–[61]. In RIR, methods usually retrieve information in the spatial domain, as it usually provides more flexibility regarding the type of cues that can be used.

3.1.2 Global vs local methods

Methods based on similarity of intensities are referred to as global methods [26], [28], [42], [56], as they utilize the entirety of image pixels. In retinal images, these methods are usually based on mutual information [26], [56]. Instead of employing all image pixels, certain feature-based methods rely on carefully selected, localized features. Feature-based approaches are known as local methods [6], [9], [11], [12], [23]–[25], [27], [29]–[32], [36], [37], [43]–[55], [57]–[61].

Feature based methods are preferred for registering image pairs with a small overlap. These pairs exhibit an increased registration difficulty, due to the small amount of commonly available information. They are also preferred for registering images with anatomical changes, as features are robust to local image differences. As such, they are stronger cues to perform registration between images, while in general, they require less processing power, leading to faster registration.

3.1.3 Transformation model used

Image registration can use different transformation models to warp the target image to the reference image. Such transformations typically include 2-D (linear, affine) or 3-D (projective, quadratic) transformations. Linear transformations consist of rotation, scaling, and translation [23], [28], [29], [36], [37], [42], [43], [45], [49], [53], [56], [57], [59]. Affine transformations preserve straight and parallel lines, but not angles [9], [12], [23]–[29], [36], [37], [44]–[50], [53], [54], [57]–[60]. Projective transformations perform a perspective distortion [6], [11], [27], [30]–[32], [44], [53]. Quadratic transformations allow for a more flexible warping of the test image [9], [23], [27]–[32], [36], [37], [45], [47], [49], [51]–[53], [55], [58], [59], [61].

For retinal images with a narrow Field of View (FOV), low order transformations such as linear, affine or projective may be adequate. However current image acquisition devices allow for obtaining images with larger FOV, in which the eye surface in the periphery of the images appears distorted. For these cases, the high order warping provided by quadratic transformations proves to be more suitable.

3.1.4 Intra-modal vs cross-modal image registration

This classification differentiates between methods that register images obtained using the same technique, or the same kind of sensor (intra-modal) [6], [24], [25], [28], [30]–[32], [42], [43], [47], [50], [52]–[54], [57], [59], [61] and methods that also register images obtained using different scanner types (cross-modal) [9], [11], [12], [23], [26], [27], [29], [36], [37], [44]–[46], [48], [49], [51], [55], [56], [58], [60]. Cross-modal registration is useful for making possible the analysis and comparison of images that may emphasize particular characteristics of the retina.

3.2 Related work

The only RIR method based on the frequency domain is [42], which is slow, only allows for 2D transformations and has a high possibility of getting stuck in local minima. In the RIR case, spatial methods seem to be a better alternative.

Within spatial methods, several utilize 2D transformations [12], [24]–[26], [43], [46], [48], [50], [54], [56]–[58], [60], or projective transformations [11], [44]. They provide accurate registration results for images with narrow FOV or large overlap, but their performance decreases otherwise.

Methods based on quadratic transformations are deemed to be the ones with the potential to achieve the most accurate registration. Within methods that use quadratic transformations, global ones such as [28] are computationally expensive due to the large amount of information used. Local methods are typically faster to execute. Speeded Up Robust Features (SURF) and Partial Intensity Invariant Feature Descriptor (PIIFD) based keypoints such as [29], [37], [55], [61] are shown [62] to provide less accurate registration results than other keypoint features such as Scale-Invariant Feature Transform (SIFT) [31], [32], [62] and vessel bifurcations [9], [23], [27], [47], [51], [52], [62]. A series of closely related methods based on Iterative Closest Point (ICP) [36], [45], [49] provide accurate registration results when certain conditions are met. However, they contain an initialization step based upon a single SIFT keypoint match, which proves to be a weakness in challenging image pairs [32]. RIR registration methods typically do not utilize simultaneously different types of features.

Methods that utilize second order polynomials for the transformation, may be able to obtain accurate registration for the overlapping areas in the images. However, the transformation performed in non-overlapping areas might not be accurate. Such a drawback could be overcome by utilizing a model approximating the shape of an eye [11], [31].

In this work, a local registration method in the spatial domain that performs registration of intra-modal retinal images utilizing several eye models is proposed. The main novelty of the proposed method consists on generating a 3D scene in which the relative pose of the cameras that acquired the images, as well as the shape of an ellipsoidal model are simultaneously estimated. Parts of this method have appeared in the literature [30]–[32], [62], but the framework is presented in more detail in Chapter 6.

Table 3.1 summarizes the classification of the reviewed RIR methods. More details about each registration method are provided in Appendix A.

Method	Freq.	Spa.	Local	Global	Intra.	Cross.	Linear	Affine	Projective	Quadratic
A.1 Peli (1987) [43]		×	×		×		×			
A.2 Cicediyan (1992) [42]	×			×	×		×			
A.3 Matsopoulos (1999) [44]		×	×		×	×		×	×	
A.4 Laliberté (2003) [23]		×	×		×	×	×	×		×
A.5 Stewart (2003) [45]		×	×		×	×	×	×		×
A.6 Ryan (2004) [9]		×	×		×	×		×		×
A.7 Matsopoulos (2004) [46]		×	×		×	×		×		
A.8 Chanwimaluang (2006) [47]		×	×		×			×		×
A.9 Choe (2006) [12], [48]		×	×		×	×		×		
A.10 Yang (2007) [49]		×	×		×	×	×	×		×
A.11 Lin (2008) [11]		×	×		×	×			×	
A.12 Chaudhry (2008) [50]		×	×		×			×		
A.13 Tsai (2010) [36]		×	×		×	×	×	×		×
A.14 Chen (2010) [37]		×	×		×	×	×	×		×
A.15 Deng (2010) [51]		×	×		×	×				×
A.16 Perez-Rovira (2011) [52]		×	×		×					×
A.17 Zheng (2011) [53]		×	×		×		×	×	×	×
A.18 Troglio (2011) [24], [54]		×	×		×			×		
A.19 Gharabaghi (2012) [25]		×	×		×			×		
A.20 Ghassabi (2013) [55]		×	×		×	×				×
A.21 Reel (2013) [26]		×		×	×	×		×		

Table 3.1: Table continues in next page. Classification of registration methods according to 4 different criteria. Frequency or spatial domain. Local or global method. Intra-modal and / or cross-modal. Type of transformation.

Method	Freq.	Spa.	Local	Global	Intra.	Cross.	Linear	Affine	Projective	Quadratic
A.22 Legg (2013) [56]		×		×	×	×	×			
A.23 Bathina (2013) [27]		×	×		×	×		×	×	×
A.24 Hernandez-Matas (2014) [6]		×	×		×				×	
A.25 Chen (2014) [57]		×	×		×		×	×		
A.26 Adal (2014) [28]		×		×	×		×	×		×
A.27 Lee (2015) [58]		×	×		×	×		×		
A.28 Wang (2015) [29]		×	×		×	×	×	×		×
A.29 Ghassabi (2015) [59]		×	×		×		×	×		×
Hernandez-Matas (2015) [30]–[32], [62]		×	×		×				×	×
A.30 Liu (2016) [60]		×	×		×	×		×		
A.31 Saha (2016) [61]		×	×		×					×

Table 3.1 continued: Classification of registration methods according to 4 different criteria. Frequency or spatial domain. Local or global method. Intra-modal and / or cross-modal. Type of transformation.

Chapter 4

Applications

Retinal Image Registration (RIR) can be utilized as a stepping stone for several applications that aim to facilitate the analysis of the retina by clinicians. Such applications range from increasing the retinal area displayed by an image, to enhancing the quality of the picture, estimating geometrical characteristics, video stabilization and even tracking changes across time such as the thinning of blood vessels. In this chapter, we give an overview of some of these applications.

4.1 Longitudinal studies

The main purpose of registering retinal images from different time periods resides in being able to monitor the evolution of retinopathy in the patient [61]. This can be used both to track the usefulness of a treatment (to be able to see if the patient recovers, and how fast), as well as to follow the evolution of the sickness in untreated patients. Ways to do this would be to precisely compare variations in vessel diameter at the same anatomical points, or to observe the growth of cotton-wool spots.

There exist plenty of methods to detect general changes in image sequences [63], but there is not much work done on the detection of changes in retinal images. Tracking pixel intensity changes (for color fundus images), as the methods described below, may prove useful if those changes can be identified and classified. However, those methods are not effective for detecting early arteriolosclerosis, which is the most important symptom for early diagnosis of hypertension.

Narasimha-Iyer et al. (2005) [15] register the images using Generalized Dual Bootstrap-Iterative Closest Point (GDB-ICP). Then they correct illumination and perform change detection and classification using Bayesian classifiers.

Troglio et al. (2009) [16], [24], [54] register the images using genetic algorithms. Then they calculate a difference map. Contrary to Narasimha-Iyer et al. they don't perform any classification of the data from the difference map.

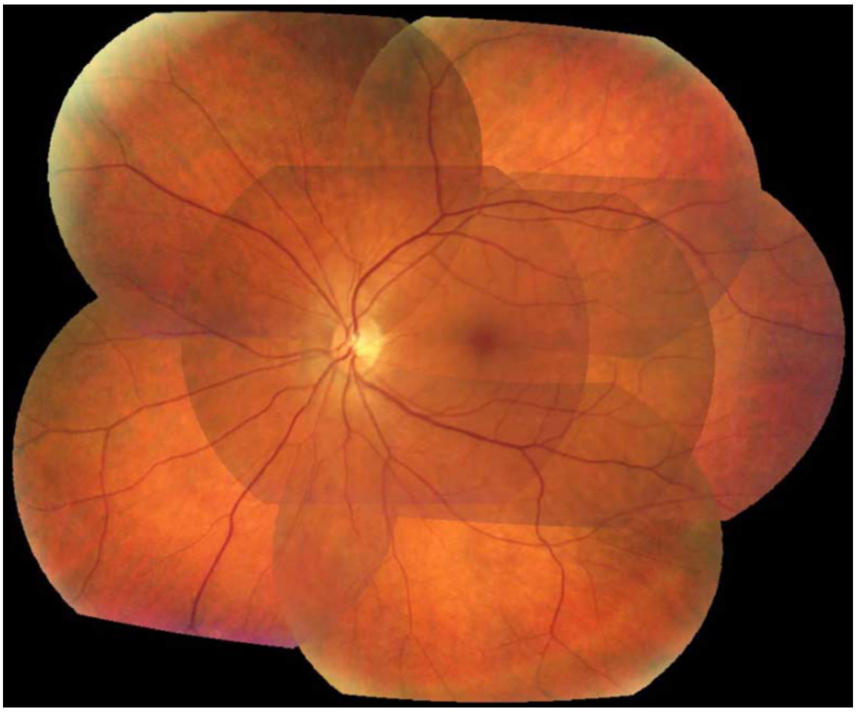


Figure 4.1: Mosaic image composed by several retinal images [9].

4.2 Mosaicing

Traditionally, RIR has been employed mostly to perform image mosaicing. This practice consists of aligning retinal images from different parts of the retina to create a representation with a wider Field of View (FOV). Typically, mosaicing is performed with images acquired during the same examination session that image areas of the retina with small overlap.

Old fundus cameras generally have a narrow FOV, thus the importance of mosaicing to properly study the retina. New cameras keep appearing, increasing their FOV. This trend indicates that in the future mosaicing will become less relevant. However, presently this is still the most important application of RIR. Figure 4.1 depicts and example of mosaicing extracted from [9].

4.3 Super resolution

Multi-frame Super Resolution (SR) methods utilize multiple images of the same scene acquired from slightly different viewpoints to provide an image of higher resolution and definition [64]. In fundoscopy, imaging the retina from slightly different perspectives, even when attempting to image the same surface, is inherent due to saccadic motion. The basis of SR methods is image registration, because it enables the utilization of pixels from different images to be considered as additional samplings of the same function.

The general model of the SR problem states that a given image captured by a device, is a noisy, blurred and down-sampled representation of a scene. Via a combination of multiple images of the same scene, a closer to reality version of the scene can be reconstructed in a SR image [64]. The prerequisite for such an operation is a registration mapping across the input images, in order for the multiple samples to be spatially combined. The resulting image is provided at a higher resolution than the original. The increment in resolution is called the scaling factor. It is worth noting that if the scaling factor is n, and l is the number of images, if $l < n^2$ the problem is undetermined, if $l = n^2$ it is a square problem, and if $l > n^2$ the problem is overdetermined [64].

Once the warping between the reference and the rest of the images has been determined, conventional SR methods can be applied. That is, via SR algorithms, we can produce images that are closer to reality than the original images employed. This is important for retinal images because it enables more precise measurements, like the ones required for techniques such as the Arteriolar-to-Venular Ratio.

Most SR methods treat the problem of combining multiple images in a 2D context [65]–[68]. This has been also the case in the application of SR methods in retinal images [4], [5], [69]. That is, the underlying registration method considers only 2D translation and rotation, assuming that perspective differences across images are negligible. Although there exist RIR methods that perform 3-D warping of the retinal images [6], [9], [11], [12], [23], [27]–[32], [36], [37], [44], [45], [47]–[49], [51]–[53], [55], [58], [59], [61], to the best of our knowledge, only [6], [31] have been employed in the problem of SR for retinal images. Not related to retinal imaging, the work in [70] considers surface orientation and camera pose but only for 3D surface patches rather than entire surfaces.

Meitav et al. (2011) [4] proposed a super resolution method that performs cross correlation between the reference image and the image to register, finds the maximum correlation and performs 'shift and add' frame with correlation weight. This process is repeated iteratively for the rest of images.

Molodij et al. (2014) [5] perform frame selection using a contrast criterion, then shift and rotate the images to register using the Weber-Fechner criterion to select the best result. A quality assessment is performed to select the best percentage of images. Finally, they add the frames.

Köhler et al. (2014) [69] perform photometric registration, then geometric registration and finally they apply the SR method devised by Farsiu et al [64].

4.4 Retina and vessel tree reconstruction

Registered retinal images can be used to perform 3D reconstructions of the retina and / or its vessel trees. As with SR, these reconstructions and estimations can assist clinicians in the form of more precise measurements of diverse elements in the retina. There exist several different approaches to these reconstructions. While some attempt to reconstruct the retinal surface [11], [48], [71], [72], others focus on reconstructing the 3D structure of the vessel tree [73], [74].

4.4.1 Retinal surface reconstruction

Laliberté et al. (2005) [71] detect vessel bifurcations, and then calculate the fundamental matrix that describes the epipolar constraint. This matrix is used to calculate a dense disparity map using energy maximization. This is done via either partial differential equations with scale space, or using maximum-flow. It is worth noting that the surface of the retina is not really reconstructed, but a disparity map of the distribution of the fluorescein within the eye is calculated.

Choe et al. (2006) [12], [48] detect vessel bifurcations, and then estimate the fundamental matrix using plane+parallax algorithm from Kumar et al. [75]. This matrix is utilized to perform image rectification. Next, they perform stereo matching via mutual information. Finally, they calculate the disparity map.

Lin et al. (2008) [11] extract Scale-Invariant Feature Transform (SIFT) features on edge response images and match them via iterative nearest neighbor. Then they compute chained registration of the images. They estimate the fundamental matrix using plane+parallax method [75], perform a 4-pass bundle adjustment to estimate camera pose, 3D structure and refine them. Then they compute dense correspondences between the images and a disparity map using window based stereo matching and mutual information. Finally, they register 3D images using back projection.

Cheriyan et al. (2012) [72] propose a 3D retinal reconstruction method using single images (in contrast to the usual pair of images). It is based on Shape From Shading. This method uses a linear approximation and Lambertian reflectance, as it is the most common among real world images. The steps followed are to equalize the image histogram, to calculate shape from shading using the Tsai-Shah algorithm and to perform surface rendering.

4.4.2 3D vessel tree reconstruction

Matinez-Perez et al. (2004) [73] perform camera calibration computing the fundamental matrix using the Kruppa equations [76]. Then they segment the blood vessels. Afterwards they compute projection matrices from the fundamental matrix. Finally, they compute the 3D point that may create the projection of the point in both images by triangulation.

Liu et al. (2009) [74] extract vessel bifurcations and match them. Then they estimate the epipolar geometry, and they compute the fundamental matrix. Next, they calculate the projection matrix. Finally, they reconstruct the 3D points using an Iterative-Eigen linear method.

4.5 Video stabilization

Registering the frames of a video from the retina [77] to a single frame allows to create a new video in which the retina remains stable. A stable video permits a more accurate observation of the blood flow through the vessels. These same frames can be employed in creating and analyzing blood flow models. In the literature, it has also been utilized to compensate for retinal motion and hence enhance the accuracy, speed, and patient safety of retinal laser treatments [77].

4.6 Denoising

To reduce the impact of some of the transient components and image noise, image processing can be applied to a sequence of registered images, such as a video recording, to obtain a clearer image [78]. This application is useful for clinicians because they facilitate tasks such as counting retinal cells (quantification) or analyzing their shapes. There are several techniques to perform denoising that can be utilized after frame registration, combining the information provided by the corresponding pixels in the registered images. Such techniques are averaging, utilizing a temporal median filter [79] or Least Median of Squares [80], to mention a few.

Part II

Technical contributions

Chapter 5

Features, computational tools and models

This chapter introduces some basic blocks upon which the research presented in this work is built. These blocks range from feature extraction and matching, to optimization methods and geometrical models.

5.1 Keypoint extraction and matching

Keypoint detection and descriptor extraction methods have been widely utilized to identify common points in a pair of images. In this work, the four most widely utilized methods for Retinal Image Registration (RIR) have been studied. These features are Scale-Invariant Feature Transform (SIFT) [81], Speeded Up Robust Features (SURF) [82], Partial Intensity Invariant Feature Descriptor (PIIFD) on Harris corners [37] and vessel bifurcations.

5.1.1 SIFT

SIFT [81] 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. RIR methods utilizing SIFT keypoints are found in [11], [31], [32], [36], [49], [55].

SIFT chooses keypoint locations at maxima and minima of a Difference of Gaussian (DoG) function applied to an image pyramid with smoothed and resampled images. Edge response points along an edge, as well as points with low contrast are discarded. For each location, image gradients and orientations are extracted from each level of the image pyramid. The most often occurring gradient is selected as the canonical orientation.

For each keypoint location, image gradients and orientations around the keypoint are inserted into 8 orientation bins in 4×4 descriptors computed from a 16×16 sample array. This creates a descriptor consisting on a vector with 128 elements, which values are then normalized to unit length.

5.1.2 SURF

SURF [82] is another widely used method to detect and represent keypoints. It is partially invariant to image translation, scaling, rotation and illumination. RIR methods employing SURF keypoints are found in [6], [29], [30], [61].

Box filters approximating Gaussian second order derivatives are used in integral images. Then, a Hessian matrix-based blob detector is utilized in several scales to select points of interest. Keypoints are at locations where the determinant has a maximum value. Due to the utilization of box filters and integral images, scale space is not implemented as an image pyramid, as is the case for SIFT. Instead the filter size is up-scaled. This leads to potentially faster detection of keypoint locations compared to SIFT.

Haar-wavelet responses in x and y directions are utilized in the neighborhood of the interest point to calculate the point orientation. A square region around the point is extracted and divided into 4×4 square sub-regions. For each sub-region, a 4-dimensional descriptor vector is calculated via Haar wavelet responses at 5×5 sample points. A SURF descriptor vector is 64 elements long.

5.1.3 Harris-PIIFD

PIIFD on Harris corners (Harris-PIIFD) [37] is a method introduced with the purpose of finding keypoints in cross-modal retinal image pairs. PIIFD is invariant to image rotation, and partially invariant to image intensity and perspective change. RIR methods employing PIIFD are found in [29], [37], [55].

A Harris corner detector is utilized to find distinctive points in the image. To extract PIIFD, image gradient magnitudes and orientations are sampled in the local neighbourhood of the corner point. A square region following the main orientation around the point is extracted and divided into 4×4 square sub-regions. Gradient magnitudes are normalized piecewise, and the 16-bin orientation histogram is converted to an 8 bin histogram by adding the bins in opposite directions. The descriptor consist of a linear combination of the subregions histograms and their $180°$ rotated version. PIIFD is a 128 elements vector.

5.1.4 Vessel bifurcations

Vessel bifurcations, or Y-features, are commonly used in RIR [9], [12], [23], [45]–[47], [50], [57]. In this work, bifurcations are extracted conventionally as in [50]. For each image from the pair, the vessel tree of the retina is segmented utilizing Frangi's multiscale vessel enhancement filtering [83], which provides a grayscale image indicating the vesselness of each pixel. A binary threshold is applied, and the resulting binary image is skeletonized using Zhang's thinning method [84]. On the skeletonized image, bifurcations are calculated by selecting vessel pixels that in the 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. Using the extracted descriptors, the resultant keypoints are matched conventionally as if they were SIFT keypoints.

5.1.5 Keypoint matching

Matching of features is performed with a bilateral matching based on the conventional method proposed by Lowe [81]. Keypoint detection and feature descriptor extraction is performed in a pair of images. Nearest neighbor utilizing the minimum Euclidean distance for the descriptor vector is used to find the best candidate. This is performed via best-bin-first, a modified version of the k-d tree search algorithm. All matches in which the distance ratio between the closest neighbour and the second-closest neighbor is greater than 0.8 are discarded, to remove outliers. This matching process is performed from keypoints of image A to image B and from image B to image A, and only the matches that are common to both processes are kept.

5.2 Rigid pose estimation

There exist methods that can be utilized to obtain a preliminary camera pose, such as Random Sample Consensus (RANSAC) [85] and posest [86]. A preliminary camera pose can be used as an initialization to the proposed framework, to perform more accurate camera pose and eye shape estimation.

5.2.1 Random Sample Consensus (RANSAC)

RANSAC is a framework that enables outlier detection. It was initially introduced by Fischler et al. [85] as a solution to the Perspective-n-Point (PnP) problem, which is the purpose for which it is utilized in this work. This method estimates the 3D pose of an object given a set of 2D-3D correspondences and the camera projection matrix P. The 3D pose is calculated by minimizing the reprojection error between the 2D points and the projected 3D points and it is robust to the presence of outliers.

Instead of utilizing as much data as possible to obtain a solution, and then eliminating invalid data points, RANSAC randomly samples minimal samples of available data, generating a model for each set, and retaining the model with the largest amount of inliers.

5.2.2 Posest

Posest is a framework for performing model-based pose estimation for rigid objects, introduced by Lourakis and Zabulis [86]. First, it calculates a preliminary pose estimation. For that, it uses a PnP solver embedded in a RANSAC framework, using the redescending M-estimator sample consensus (MSAC) cost function. RANSAC is used to determine the best hypothesis as well as to classify correspondences into inliers and outliers. Next, a non-linear refinement of the preliminary pose is performed iteratively with the Levenberg-Marquardt algorithm. To avoid entrapment in local minima during the non-linear refinement, multi-start global optimization, using multiple local optimization processes initialized at different starting points, can be utilized.

5.3 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) was originally introduced in 1995 by Kennedy et al. [87] but it has evolved across time, with one of its most relevant snapshots being described by Poli et al in 2007 [88]. PSO achieves optimization by iteratively improving a candidate solution given an objective function. The method solves the problem by utilizing 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.

An array of particles is initialized with random positions $\vec{x}_i$ and velocities $\vec{v}_i$ across the search space. After this initialization, the following process is performed iteratively. The objective function is evaluated in each particle's location. Each particle is compared with its own best evaluation, and if it performed better in the current iteration, the particle's best location $\vec{p}_i$ is updated as $\vec{p}_i = \vec{x}_i$. The overall best performing particle is identified and its index assigned to the variable k. Then $\vec{v}_i$ and $\vec{x}_i$ are updated as follows:

$$\vec{v}_i = \vec{v}_i + \vec{U}(0, \phi_1) \cdot (\vec{p}_i - \vec{x}_i) + \vec{U}(0, \phi_2) \cdot (\vec{p}_g - \vec{x}_i), \tag{5.1}$$

$$\vec{x}_i = \vec{x}_i + \vec{v}_i, \tag{5.2}$$

where $\vec{U}(0, \phi_i)$ consists of a vector of random numbers uniformly distributed in $[0, \phi_i]$ generated at each iteration and for each particle.

PSO depends on few parameters, does not require knowledge of the derivatives of the objective function and can handle cross-modal and possibly discontinuous objective functions [89]. Additionally, it requires a relatively low number $(p \cdot g)$ of objective function evaluations, which is called 'budget' of the optimization [89]. A small budget will terminate the process prematurely with a poor pose estimate, while a too large budget will lead to extra processing time without leading to noticeable improvements in accuracy. Beyond these extreme conditions, the selection of budget offers a trade-off between the accuracy and the speed of the method [89]. Additionally, for a given budget, the distribution $p \;/\; g$ of p particles and g generations is relevant to the final performance of the method.

5.4 World model

Several tools presented in this work are based on a 3D model of the scene composed by the virtual acquisition devices and the eye model. When a retinal image is acquired, typically a headrest is utilized to fix the head in place and align the camera, centering it across the eye. Such scene can be generated from a few parameters:

- Lens to cornea distance (ltc): Known from the acquisition device's manual.
- Field of View (fov): Known from the acquisition device's manual.
- Radius of the image in pixels (r): Is calculated from the utilized image.
- Radius of the eye model (ρ): Value $\rho = 12\,mm$ is chosen as per the Navarro eye model [90], and it is used as the baseline ρ.
- Camera pixel spacing in mm (p): As it will be shown, this number can be estimated and it is not really needed. In this work, $p = 0.001$ is utilized.

Focal distance is calculated as follows:

$$f = p\; r\; \frac{ltc + \rho + \rho\;\cos\left(\dfrac{fov}{2}\right)}{\rho\;\sin\left(\dfrac{fov}{2}\right)}. \tag{5.3}$$

A graphical representation is shown in Figure 5.1. This world geometry is the cornerstone of several of the tools presented in this research, as it allows to transfer a fundus image from its 2D form to the real world, and to project it into a 3D eye model and vice versa.

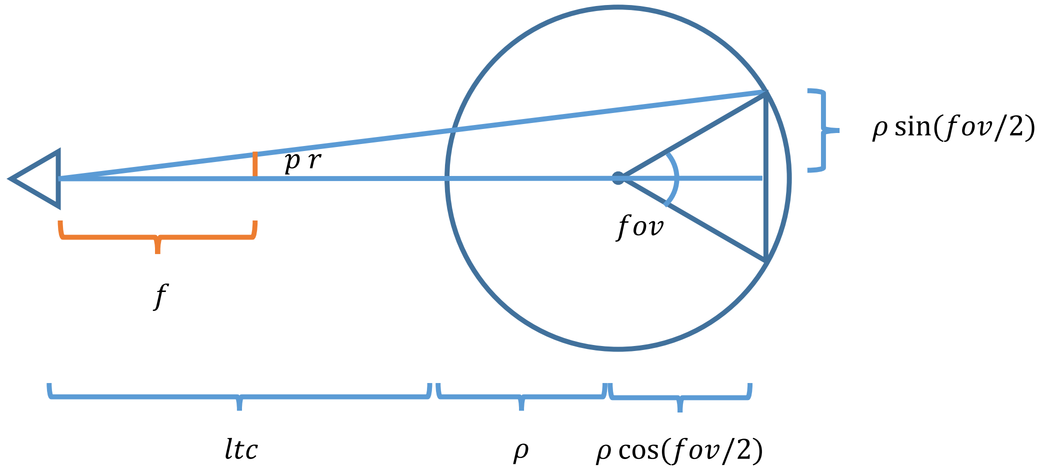


Figure 5.1: World geometry.

5.5 Eye models

The general shape of the eye can be represented via different geometrical models. In this work, four distinct eye models are utilized. These models are a plane, a sphere, an ellipsoid with fixed pose, and an ellipsoid with rotational freedom. All models, including the sphere and the plane, can be modelled through an ellipsoidal model $\{\mathbf{A}, \mathbf{Q}\}$. Locating the coordinate origin at the center of the model surface $\mathcal{E}$, we obtain the ellipsoid equation

$$\mathbf{x}^T\mathbf{Q}^T\mathbf{A}\mathbf{Q}\mathbf{x} = 1, \tag{5.4}$$

where $\mathbf{x}$ is a point on $\mathcal{E}$. This model has 3 orthogonal semi-axes $[a, b, c]$ composing $\mathbf{A}$ as shown in Equation 5.5

$$\mathbf{A} = \begin{bmatrix} a^{-2} & 0 & 0 \\ 0 & b^{-2} & 0 \\ 0 & 0 & c^{-2} \end{bmatrix}, \tag{5.5}$$

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.

5.6 Parameterization of eye shape and camera pose

A parameterization of eye shape and camera pose $\mathbf{S}$ in the world model consists of relative camera pose $\{\mathbf{R}, \mathbf{t}\}$ and ellipsoid model $\{\mathbf{A}, \mathbf{Q}\}$, thus $\mathbf{S} = \{\mathbf{R}, \mathbf{t}, \mathbf{A}, \mathbf{Q}\}$, totalling 12 parameters. Relative camera pose $\{\mathbf{R}, \mathbf{t}\}$ corresponds to the relative rotation $\mathbf{R} = \mathbf{R}_x(r_\theta) \cdot \mathbf{R}_y(r_\phi) \cdot \mathbf{R}_z(r_\omega)$ and translation $\mathbf{t} = [t_x\, t_y\, t_z]^T$ between the cameras. Ellipsoidal model $\{\mathbf{A}, \mathbf{Q}\}$ is described in Section 5.5.

5.7 Intrinsic camera matrix

The intrinsic camera matrix K describes the geometry of a camera, whether it is virtual or real. The matrix is defined in Equation 5.6. K, together with the extrinsic parameters $\{\mathbf{R}, \mathbf{t}\}$ of the camera allow to project 3D point positions onto a camera view.

$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.6}$$

In Equation 5.6 α_x and α_y represent the focal length in pixels, γ 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 cameras utilized in our experiments, present no skew, thus $\gamma = 0$. The central point of the image (u_0, v_0) is calculated from the fundus image utilized in the registration. α_x and α_y are calculated from f in Equation 5.3 as follows:

$$\alpha_x = \alpha_y = \frac{f}{p} = r \, \frac{ltc + \rho + \rho \ \cos\left(\dfrac{fov}{2}\right)}{\rho \ \sin\left(\dfrac{fov}{2}\right)}. \tag{5.7}$$

5.8 Ray-ellipsoid intersection

Within the world model, given a parameterization $\mathbf{S}$, the points in a fundus image can be traced to 3D locations on the surface $\mathcal{E}$ of model $\{\mathbf{A}, \mathbf{Q}\}$. This is performed by calculating the line traversing from the camera optical center to the keypoint in the image. The intersection of the line and $\mathcal{E}$ indicates the 3D position of the point $\mathbf{x}$. To estimate $\mathbf{x}$, the line equation from the camera center $\mathbf{c}$ and through pixel $\mathbf{u}$ is solved for λ:

$$\mathbf{x} = P^{+}\mathbf{u} + \lambda\mathbf{c}, \tag{5.8}$$

$$P^{+} = P^{T}(PP^{T})^{-1}, \tag{5.9}$$

as per Equation 6.13, [91, p. 162].

Chapter 6

REMPE registration framework

In retinal imaging, the acquisition device typically has a fixed pose, while it is the eye that rotates in place, allowing to image diverse areas of the retina. In this work, however, an equivalent approach is followed. In it, the eye is fixed in place, and the acquisition device may image it from diverse poses.

The proposed framework, named Registration through Eye Modelling and Pose Estimation (REMPE) and publicly available at `http://www.ics.forth.gr/cvrl/rempe`, is shown in Figure 6.1. It registers the reference (F_0) and test (F_t) images by simultaneously estimating the two relative poses of the acquisition device, as well as the 3D shape and 3D orientation of an ellipsoidal eye model. The eye model has semi-axes $[a, b, c]$ and rotations along said semi-axes $[r_a, r_b, r_c]$ leading to eye model surface $\mathcal{E}$. The eye model is centered at $\mathbf{c}_s = [0, 0, 0]^T$. 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. Point correspondences between the images are utilized to achieve this registration. An initial pose estimate is calculated utilizing Random Sample Consensus (RANSAC) and a spherical eye model. Subsequently, Particle Swarm Optimization (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 Figure 6.2.

6.1 Keypoint correspondences

During the development of this research work, preliminary experiments regarding the type of cues to utilize were performed. Those experiments involved global cues such as Sum of Absolute Differences or Cross Correlation, and local features like vessel trees, or distance transforms upon these trees. However, these preliminary experiments showed that utilizing keypoint correspondences provided more accurate results, at a much lower computational cost. Due to this, in this work the emphasis was set on keypoint correspondences.

The impact of Scale-Invariant Feature Transform (SIFT), Speeded Up Robust

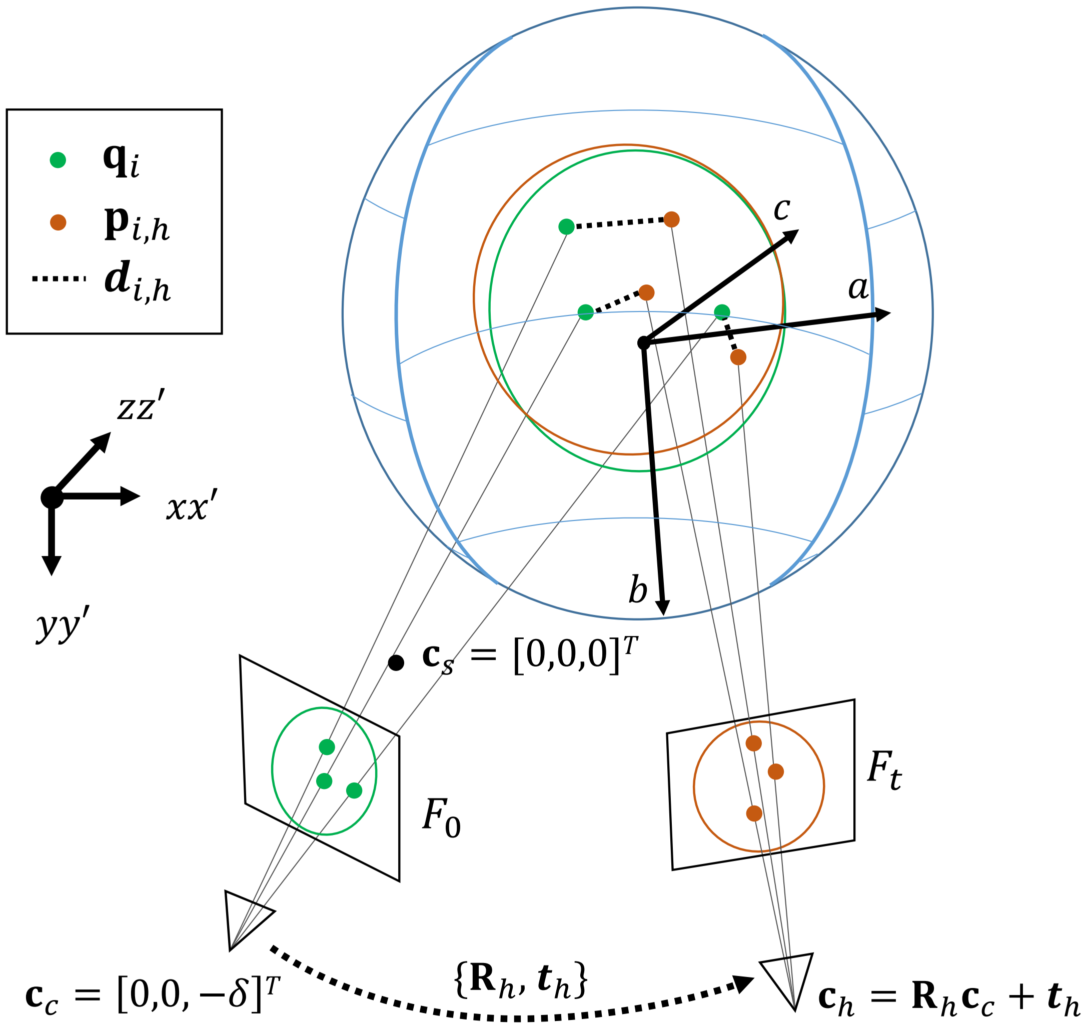


Figure 6.1: Geometry of the proposed registration framework.

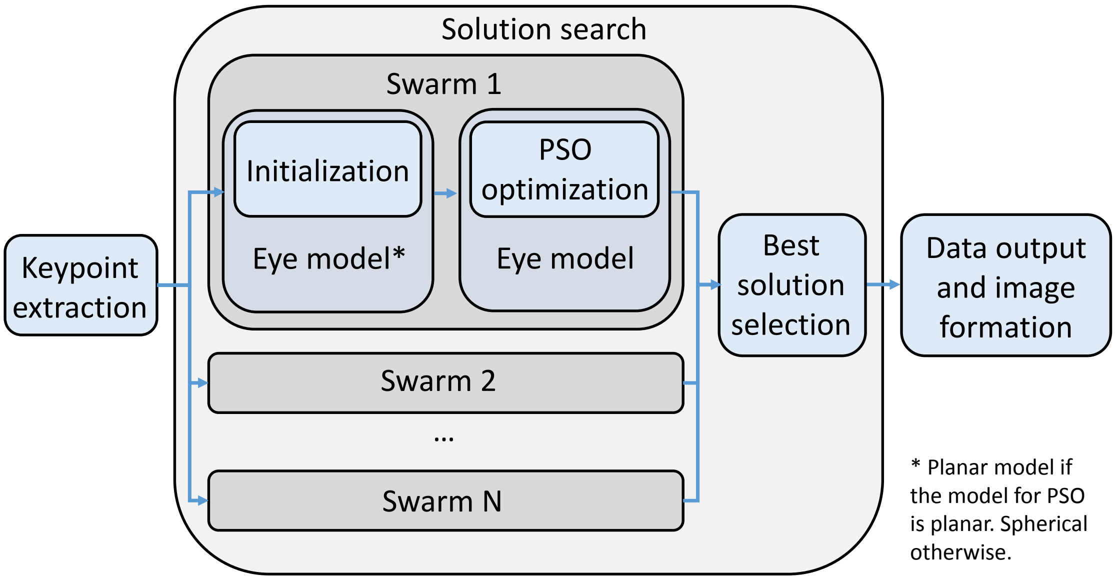


Figure 6.2: Workflow of the proposed registration framework.

Features (SURF), Partial Intensity Invariant Feature Descriptor (PIIFD) on Harris corners and bifurcations as keypoints used by the proposed retinal image registration framework is studied. These keypoints have been selected for being the most widely employed for retinal image registration. These keypoints are reviewed in more detail in Section 5.1.

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.

For the proposed method the combination of SIFT keypoints and bifurcations is utilized, as the experiment in Section 8.2 identify this combination as the most accurate and robust.

6.2 Initialization

An initial candidate solution $\mathbf{S}_0$ is estimated by solving the Perspective-n-Point (PnP) problem. Two different approaches were considered. The first approach is RANSAC [85], which calculates $\{\mathbf{R}, \mathbf{t}\}$ through the minimization of the projection error between 2D and the 3D corresponding points. RANSAC is discussed in more detail in Section 5.2.1. The second approach is posest [86]. Posest performs a preliminary pose estimation, followed by a non-linear refinement. Multi-start global optimization is utilized to avoid entrapment in local minima. Further details are provided in Section 5.2.2.

Both methods rely on 2D-3D correspondences. As the matched keypoints provide us only with 2D-2D correspondences, and we have no knowledge regarding $\{\mathbf{A}, \mathbf{Q}\}$, we utilize the spherical model and the calibration of the cameras to retrieve the 3D location of the points utilizing Equation 5.8.

If no initialization method is used, $\{\mathbf{R}, \mathbf{t}\}$ is initialized as $\{\mathbf{I}, \mathbf{0}\}$.

6.3 Optimization

The estimation of $\mathbf{S}$, is formulated as the solution of an optimization problem whose goal is the minimization of an objective function.

Given $\mathbf{S}_h$ with $\{\mathbf{R}_h, \mathbf{t}_h\}$, the keypoints of the image are traced to 3D locations on the eye model surface $\mathcal{E}$. This is performed by calculating the line traversing from the camera optical center to the keypoint in the image. The intersection of the line and $\mathcal{E}$ indicates the 3D position of the point, and it is calculated utilizing Equation 5.8.

Points $\mathbf{q}_i$ 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 $\mathbf{S}_h$ (see Figure 6.1). The 3D distances of corresponding keypoints on $\mathcal{E}$ are:

$$d_{i,h} = |\mathbf{q}_i - \mathbf{p}_{i,h}|. \tag{6.1}$$

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

$$o(\mathbf{S}_h) = \sum_j d_{j,h}, \tag{6.2}$$

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

6.4 Particle Swarm Optimization (PSO)

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

$$[\gamma_\theta, \gamma_\phi, \gamma_\omega, \mu_x, \mu_y, \mu_z, \mu_a, \mu_b, \mu_c, \gamma_a, \gamma_b, \gamma_c], \tag{6.3}$$

for parameters

$$[r_\theta, r_\phi, r_\omega, t_x, t_y, t_z, a, b, c, r_a, r_b, r_c], \tag{6.4}$$

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 Equation 6.2 is computationally prohibitive. Besides, that would create a discretized version of the problem. Thus, the objective function of Equation 6.2 is optimized via PSO [88], a stochastic, derivative-free optimization method successfully employed for pose estimation not only in this [30], [31] but also in other domains [92], [93]. PSO is discussed in more detail in Section 5.3.

6.5 Search space structure variants

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

- **Coarse (C):** Baseline method described in Section 6.4.
- **Coarse-to-Fine (CF):** PSO is executed two times. In the second execution, the search space is a smaller hypercube centered at the estimate provided by the first execution. The same budget is utilized in both executions, splitting the total budget in half. The hypercube is reduced in all 12 dimensions thus performing a denser search. This variant is introduced to study if performing an initial broad search, followed by a denser one around the initial estimation provices more accurate registration results.
- **RANSAC (R):** In this approach, $\{\mathbf{R}_0, \mathbf{t}_0\}$ constitutes the solution, i.e., the RANSAC-based initialization described in Section 6.2.
- **R-C:** Coarse variant with RANSAC initialization.
- **R-F:** Fine variant with RANSAC initialization. Similar to variant R-C, but with reduced search space.
- **R-CF:** Coarse-to-Fine with RANSAC initialization.

Thus, variants C and CF are initialized at $\{\mathbf{I}, \mathbf{0}\}$. Variants R-C and R-CF are identical to them, but initialized at $\{\mathbf{R}_0, \mathbf{t}_0\}$.

The search hypercube in Equation 6.3, is set to

$$[4, 4, 4, 4, 4, 4, 4, 8, 8, 8, 360, 360, 360] \tag{6.5}$$

for coarse PSO processes, and to

$$[2, 2, 2, 2, 2, 2, 4, 4, 4, 180, 180, 180] \tag{6.6}$$

for fine PSO processes.

The results presented and discussed in Section 8.5 lead to selecting variant R-F as the proposed variant.

6.6 Eye model variants

In this work, four eye models to approximate either the full shape of the eye, or part of it, are utilized. All of them present a smooth surface. While these models do not accurately represent the retinal surface and structure in fine detail, they establish the ground work for attaining an accurate estimation. In increasing order of complexity, the models utilized are the following:

- **Plane:** The baseline model of our method. It consists of modelling the retinal surface as a plane. While this may be appropriate for fundus images with a reduced Field of View (FOV), this approximation deviates when images with large FOV are used. $[a,\, b,\, c] = [10000, 10000, 1]$ and $[r_a,\, r_b,\, r_c] = [0, 0, 0]$.

- **Sphere:** The spherical model is the simplest full eye approximation. The sphere consists of an ellipsoid with equal axes, which also allows for ignoring their rotation. Thus $[a,\, b,\, c] = [\rho, \rho, \rho]$ and $[r_a,\, r_b,\, r_c] = [0, 0, 0]$. Value $\rho = 12\, mm$ is chosen as per the Navarro eye model [20], [90], [94].

- **Fixed orientation ellipsoid:** An ellipsoid has three major axes of different lengths. The fixed orientation ellipsoidal model has a defined rotation that fixes semi-axes a, b, c to be parallel to the x, y and z axes respectively. Thus $[a,\, b,\, c] = [\rho_a, \rho_b, \rho_c]$ and $[r_a,\, r_b,\, r_c] = [0, 0, 0]$.

- **Ellipsoid:** The ellipsoidal model has three major axes of different lengths. These axes can exhibit any rotation with relation 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 order of complexity, the higher the capability of the model to correctly approximate an actual eye shape, leading to more accurate registration results.

6.7 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 gave 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 not as effective in avoiding these local minima entrapment.

6.8 Simultaneous registration of multiple images

The proposed framework allows for simultaneously registering multiple test images F_{t_i} to F_0. This property of the framework is mainly useful for the process of eye shape estimation. When a pair of images F_0 and F_t is utilized, the information utilized for estimating the relative camera pose between the images, and more importantly, the eye shape, is $F_0 \cap F_t$. Given a set of test images $\{F_{t_1}, F_{t_2}, \cdots\}$, the useful retinal information utilized is $(F_0 \cap F_{t_1}) \cup (F_0 \cap F_{t_2}) \cup \cdots$, potentially allowing the utilization of larger surface of the retina to estimate its shape.

This simultaneous registration is achieved through a modification of PSO. For a given generation g_i in the PSO process, all particles p_{i_1} for registering F_0 and F_{t_1} are run. Then, within the same generation, all particles p_{i_2} for registering F_0 and F_{t_2} are run, with the peculiarity that while a different $\{\mathbf{R}_i,\ \mathbf{t}_i\}$ is estimated, the particles lie in the same $\{\mathbf{A},\ \mathbf{Q}\}$ locations as the corresponding particles for the registration of F_0 and F_{t_1}.

This implementation allows to simultaneously register any number of F_{t_i} to a single F_0, providing a $\mathbf{S}_i$ for each pair. While registration of a single F_t to F_0 has a search space of 12 dimensions, broken down into 6 for $\{\mathbf{R}, \mathbf{t}\}$ and 6 for $\{\mathbf{A}, \mathbf{Q}\}$. Each subsequent extra F_{t_i} adds 6 dimensions to the search space, 6 for each extra $\{\mathbf{R}_i, \mathbf{t}_i\}$. Thus, simultaneously registering two F_t to F_0 has a search space of 18 dimensions, and three F_t leads to 24 dimensions, subsequently increasing the size of the search space for the solution.

An alternative approach is to perform the pairwise registrations independently, and utilize bundle adjustment [95] to simultaneously refine all parameters at a final step. However, the utilized approach provides accurate results for this stage of the research.

6.9 Data output and image formation

Data can be output in several ways. When the solution $\mathbf{S}$ is chosen, Equation 5.8 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 a list of points with the 3D coordinates and the color data 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 reconstruct the image utilizing bilinear interpolation, which is then output as an image file.

Additionally, and by default, a text file with the **S** parameters is generated, allowing to quickly generate again any of the output data at a later time.

6.10 Discussion

In this chapter, a Retinal Image Registration (RIR) framework for fundus images is proposed. The framework, which is publicly available at `http://www.ics.forth.gr/cvrl/rempe`, is based on the simultaneous estimation of the relative 3D pose of the cameras that acquired the images to be registered and the parameters and the orientation of an ellipsoidal model of the eye. To perform this process, the combination of SIFT and bifurcation matches is utilized, as combining these types of features provides increased accuracy and reliability. Keypoints are employed with a spherical eye model to perform an initial camera pose estimation using RANSAC. 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 is allows for more accurate registration than alternative polynomial transformations.

Chapter 7

Datasets

Several datasets haven been utilized in this work, some being comprised by synthetic data in either 2D or 3D and others by real images. Additionally, the tool used to generate the synthetic dataset can be utilized to map 2D retinal images to a 3D eye model.

7.1 Related work

Numerous publicly available datasets exist containing retinal images. They have been compiled for different purposes, such as segmentation of diverse elements of the retina (CHASEDB1 [96], [97], DRIONS-DB [98], [99], Drishti-GS [100], [101], DRIVE [102], [103], HRF [104], [105], MESSIDOR [106], [107], ONHSD [108], [109] and REVIEW [110], [111]), diagnosis (DIARETDB0 [112], [113], DIARETDB1 [114], [115], e-ophtha [116], [117], INSPIRE-AVR [118], [119], ROC [120], [121], STARE [122], [123] and VICAVR [124], [125]), user authentication (VARIA [126], [127]) and retinal image registration (Fundus Image Registration (FIRE) [128], [129] and RODREP [130], [131]). Table 7.1 shows a list with technical characteristics pertaining these datasets. Besides e-ophtha, FIRE, VARIA and RODREP, the rest of the datasets are not useful for the purpose of performing Retinal Image Registration (RIR) as they do not provide pairs of retinal images.

The RODREP dataset provides a large amount of images retrieved following a screening program in which for 140 eyes, 2 sets of 4 pictures are acquired. In these sets of 4 images, there is very limited overlap, as the purpose is to generate mosaics from the images. Given the characteristics of the screening, we estimate that for each eye there are approximately 10 image pairs: 3 pairs from the same session for each of the 2 sessions and 4 pairs with an image from each session. This totals 1400 image pairs in the dataset. While the amount of images and possible image pairs is impressive, we consider the dataset lacking in two respects, as far as image registration is concerned. First, there are no image pairs in which, due to progression or remission of illness, large anatomical differences exist in the retina.

Dataset	# Images	FOV	Resolution	Registrable image pairs
CHASEDB1 [96], [97]	14	$\approx 25°$	999×960	0
DIARETDB0 [112], [113]	130	$50°$	1500×1152	0
DIARETDB1 [114], [115]	89	$50°$	1500×1152	0
DRIONS-DB [98], [99]	110	$\approx 30°$	600×400	0
Drishti-GS [100], [101]	101	$\approx 25°$	2045×1752	0
DRIVE [102], [103]	40	$45°$	565×584	0
e-ophtha [116], [117]	463	$\approx 45°$	2544×1696	144
FIRE [128], [129]	129	$45°$	2912×2912	134
HRF [104], [105]	45	$45°$	3504×2336	0
INSPIRE [118], [119]	40	$30°$	2392×2048	0
MESSIDOR [106], [107]	1200	$45°$	1440×960 – 2304×1536	0
ONHSD [108], [109]	99	$45°$	640×480	0
REVIEW [110], [111]	14	$\approx 45°$	1360×1024 – 3584×2438	0
ROC [120], [121]	100	$\approx 30° - 45°$	768×576 – 1386×1391	0
RODREP [130], [131]	1120	$45°$	2000×1312	≈ 1400
STARE [122], [123]	397	$\approx 30° - 45°$	700×605	0
VARIA [126], [127]	233	$20°$	768×584	154
VICAVR [124], [125]	58	$45°$	768×584	0

Table 7.1: Publicly available retinal image datasets.

Some of these changes may correspond to microaneurysms, cotton-wool spots, hard exudates and other forms of retinopathy resulting from illnesses such as hypertension and diabetes. Registration of this kind of images is important for clinicians for the better study of retinopathy progression. Additionally, these differences may prove very challenging for automatic registration methods. Second, the dataset lacks ground truth information. Ground truth is very important for being able to evaluate quantitatively the accuracy of a registration method. The RODREP authors rely on visual inspection by medical experts to grade the registration accuracy of their registration method [130]. This is a time consuming task for medical experts, to the extent that, as reported [130], some of them did not complete the task.

e-ophtha provides several image pairs, both with large and small overlaps of the retina. But similarly to RODREP, there are no image pairs with anatomical differences, and no ground truth is provided.

VARIA also provides several image pairs. However, the Field of View (FOV) of the images is small, all the image pairs have a large overlapping area of the retina,

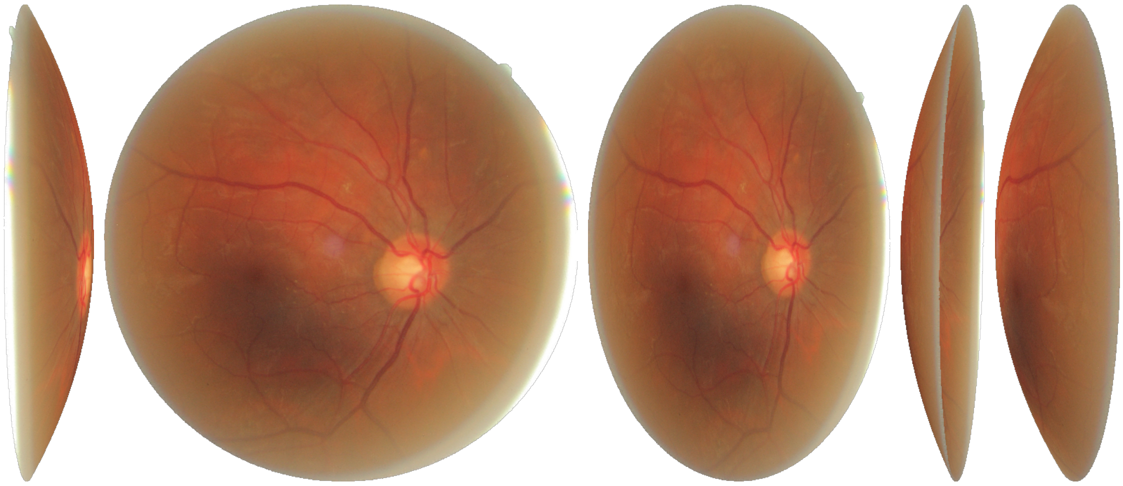


Figure 7.1: 3D synthetic data from different points of view.

and no ground truth is provided. Thus, this dataset is quite limited for the purpose of retinal image registration.

FIRE is a dataset presented as part of the work described in this dissertation, and it is covered in more detail in Section 7.3.1.

7.2 Synthetic data

Synthetic data is a very useful tool for RIR. It can be in the form of 2D images or 3D data. Synthetic 2D images can be utilized for validating registration algorithms. Synthetic 3D data can be useful for validating camera pose and eye shape estimation as well as 3D registration algorithms.

7.2.1 3D synthetic data

Given a camera pose $\{\mathbf{R}, \mathbf{t}\}$, the pixels of an image can be traced to 3D locations on the eye model surface $\mathcal{E}$ generated by $\{\mathbf{A}, \mathbf{Q}\}$. This is performed by calculating the line traversing from the camera optical center to the pixel in the image. The intersection of the line and $\mathcal{E}$ indicates the 3D position of the point, and it is calculated utilizing Equation 5.8. An example of such 3D model is shown in Figure 7.1. This allows for 3D representation of a retinal fundus image if the solution $\mathbf{S}$ is known.

7.2.2 2D synthetic data

By having a retinal texture projected into an eye model $\{\mathbf{A}, \mathbf{Q}\}$, 2D synthetic retinal images can be generated. This is done by projecting the 3D points to a given virtual camera with K parameters located in $\{\mathbf{R}, \mathbf{t}\}$. This process allows to generate 2D synthetic images with a wide range eye shapes, taken with a wide range of different

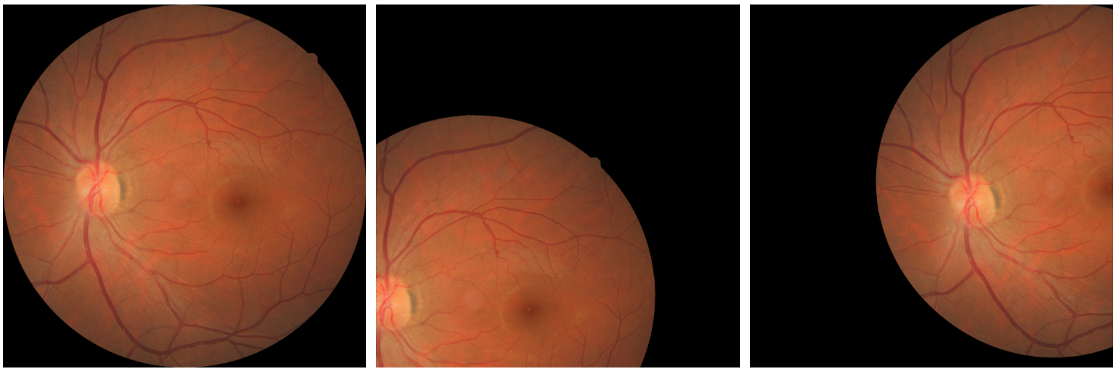


Figure 7.2: Original image (left) and 2 synthetic retinal image generated with the proposed tool (middle and right).

virtual cameras, and from a wide range of locations and orientations. Examples of such images are shown in Figure 7.2.

The advantage of generating synthetic images with this approach is multiple. Perfectly accurate control points can be generated to evaluate the accuracy of a registration method. The texture utilized to generate the image can be altered with noise or simulated retinopathy to evaluate the robustness of a registration method. The images are warped according to an eye model, so the difference between points of view in the images is realistic. Additionally, as the whole geometry and the camera parameters can be controlled the accuracy of shape estimation algorithms can be evaluated. The main drawback is that depending on the size of the original texture, and the virtual camera utilized, there may exist large areas without texture information, as shown in Figure 7.2.

7.3 Real images

For the research presented in this work, several image datasets with real images have been utilized. The main one is the FIRE dataset, which was made publicly available, and which contains three different categories $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{A}$ for different indicative registration applications. Additionally, two datasets with the aim of registration for Super Resolution (SR), under the names $\mathcal{SR}1$ and $\mathcal{SR}2$ were utilized.

7.3.1 The FIRE Dataset

The FIRE dataset [128], publicly available at `http://www.ics.forth.gr/cvrl/fire`, contains several retinal image pairs with the aim of assessing the accuracy of RIR. Two binary mask images are also included. The first distinguishes the pixels that contain color information from the ones that correspond to the black, circular frame of the image. This information is useful for global registration methods. The second mask marks a proposed area suitable for finding keypoints. This is useful for local registration methods. The images in each pair are in `.jpg` format, and are the original images acquired from the imaging device. The masks are provided in `.png` format. Additionally, ground truth point correspondences are provided for each image pair, so that they can be used to evaluate the accuracy of an image registration algorithm. These correspondences are provided in `.txt` files, one for each image pair.

FIRE consists of a collection of 129 retinal images forming 134 image pairs. For a single eye there might be several images, therefore several image pair combinations can be formed. The images were acquired with a Nidek AFC-210 fundus camera, which has a resolution of 2912×2912 pixels and a FOV of 45°× 45°. Images were acquired by Dr. Areti Triantafyllou, Dr. Panagiota Anyfanti and Dr. Stella Douma from the Hypertension Unit falling under the 3rd Department of Internal Medicine, Papageorgiou Hospital, Aristotle University of Thessaloniki, Greece, from 39 patients with an age range from 19-67 years, both males and females, who attended their regular appointments during years 2006-2015. Written informed consent was obtained before data acquisition and processing.

The image pairs are classified into three categories depending on the characteristics of the pairs, with each pair being a member of a single category:

- Category $\mathcal{S}$ (for *Similar*) 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. $\mathcal{S}$ provides images that after registration could be useful for performing super resolution.

- Category $\mathcal{P}$ (for *Pose difference*) 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. As retinal images present a relative lack of texture, that is a hurdle for keypoint extraction. These type of image pairs could be useful for mosaicing applications.

- Category $\mathcal{A}$ (for *Anatomic difference*) 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 arteriolosclerosis, 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.

Categories $\mathcal{S}$ and $\mathcal{P}$ also include pathological cases that may affect the structure of the retina, but as the images lack anatomical differences, retinopathy remains unchanged among the images of each pair. All three categories may feature eye shape deformations caused by myopia, hypermetropia and similar conditions. Category characteristics are summarized in Table 7.2.

	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$
#Image pairs	71	49	14
Approximate overlap	$> 75\%$	$< 75\%$	$> 75\%$
Anatomical changes	No	No	Yes
Indicative registration application	SR	Mosaicing	Longitudinal studies

Table 7.2: Characteristics of the FIRE dataset categories.

Figure 7.3 shows sample image pairs from the 3 categories. The leftmost column shows a pair from category $\mathcal{S}$. The second column shows a pair from category $\mathcal{P}$. In column 3, there is a pair from category $\mathcal{A}$ in which the test image exhibits anatomical differences compared to the reference image, such as increased vessel tortuosity, cotton-wool spots and hemorrhages. The rightmost column shows another pair from category $\mathcal{A}$, in which the test image exhibits additional hemorrhages, cotton-wool spots and vessel thinning, compared to the reference image.

Ground truth

In the relevant literature, the assessment of an image registration method is performed based on several quantitative and qualitative approaches. Quantitative approaches include the utilization of manually [36], [37], [45], [53], [55], [62], [132] or automatically [30]–[32] selected control points. This has the advantage of reporting registration error as a single number. Other quantitative approaches [11], [25], [26], [56] estimate the transformation (scaling, rotation and translation) that governs the solution provided by a method. A similar transformation is estimated for the ground truth solution that is approved by an expert. Then, the two transformations are compared. However, this solution usually provides the error as a collection of several parameters. Qualitative approaches usually focus on visual supervision of vessel alignment in images [130], such as counting the amount and measuring

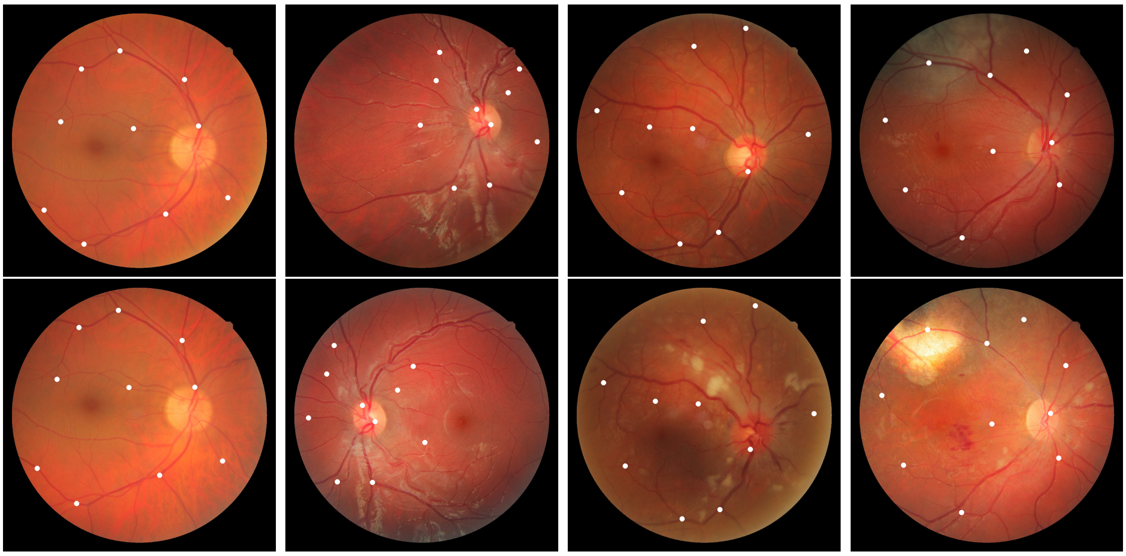


Figure 7.3: Image pairs from the FIRE dataset. Leftmost column shows a pair from Category $\mathcal{S}$, the second column a pair from Category $\mathcal{P}$ and the 2 rightmost pairs from Category $\mathcal{A}$. White dots mark control point locations.

the magnitude of discontinuities on a checkerboard superimposition after registration [52]. This task is time consuming, requires the involvement of an expert and does not allow for quantitative comparison between registration methods.

In the proposed dataset, we provide ground truth for the calculation of the registration error in the form of corresponding points between the images of the pair. These are henceforth referred to as control points. The location of a control point j in the reference image is denoted as $\mathbf{c}_j$, and the corresponding point in the test image as $\mathbf{t}_j$. A registration method is taking as input the points $\mathbf{t}_j$ and maps them to new coordinates $\mathbf{r}_j$. Thus, the $\mathbf{r}_j$ points are the $\mathbf{t}_j$ points after registration. If registration is perfect, points $\mathbf{c}_j$ and $\mathbf{r}_j$ coincide and their distance (in image pixels) is 0.

For each image pair, 10 corresponding points are provided. Correspondences are manually annotated, with about 8 of them located towards the edges of the overlapping area and the remaining ones towards the center. Points were chosen to be widespread across the image, providing a broad coverage of the overlapping surface between images, so that the accuracy across the whole image can be calculated. Points are mainly located in vessels and crossings as they were manually selected. This allowed the annotator to provide accurate initial markings, which is a challenging task with uncertain outcome in other image areas that lack image structure. The amount of correspondences was selected by balancing the trade off between time availability of the annotator, accuracy of annotations, and number of marked images.

The effect of the human error in the location of the correspondences is mitigated using computational methods as follows. For each corresponding pair of points $\mathbf{c}_j$

and $\mathbf{t}_j$, a grid of neighboring points $\mathbf{n}_{c,j}$ and $\mathbf{n}_{t,j}$ was considered. Image patches $s_{c,j}$ and $s_{t,j}$ centered around each $\mathbf{n}_{c,j}$ and $\mathbf{n}_{t,j}$ were created. The points $\mathbf{n}_{c,j}$ and $\mathbf{n}_{t,j}$ with the highest correlation between $s_{c,j}$ and $s_{t,j}$ were chosen as $\mathbf{c}_j$ and $\mathbf{t}_j$. Despite this refinement, localization error may be present in the control points. Additionally, these points do not present subpixel accuracy. In Figure 7.3, the locations of control points for four image pairs of the FIRE dataset are shown.

The refined control points are independent from widely employed automatic feature selection methods such as Scale-Invariant Feature Transform (SIFT) [81] and Speeded Up Robust Features (SURF) [82]. This prevents bias in evaluating registration methods based on such features.

7.3.2 SR datasets

Two datasets for registration and SR experiments were acquired. For each dataset, images were obtained during a single session from the same eye of the same person.

$\mathcal{SR}1$ (for *Super Resolution dataset 1*) is comprised by a subset of 9 fundus images from category $\mathcal{S}$ of the FIRE dataset.

$\mathcal{SR}2$ (for *Super Resolution dataset 2*) is the second dataset, and it contains 9 images obtained with a Heidelberg Scanning Laser Ophthalmoscope (SLO). Images have a resolution or 1536×1480 pixels and a FOV of $\approx 30°$ in each dimension. Images were provided by Dr. Maged Habib from the Sunderland Eye Infirmary, United Kingdom.

Table 7.3 summarizes the characteristics of the datasets, and reference images are shown in Figure 7.4. In both cases, the images have been resized to one third of their original size both in the x and the y axes.

Dataset	# Images	FOV	Resolution
$\mathcal{SR}1$	9	45°	2912 × 2912
$\mathcal{SR}2$	9	≈ 40°	1536 × 1480

Table 7.3: Characteristics of image pairs in $\mathcal{SR}1$ and $\mathcal{SR}2$.

7.4 Registration evaluation

Evaluation of the registration accuracy is important for several reasons. First, it shows the suitability of utilizing a given registration method. Secondly, it allows to identify weaknesses in a registration method. And finally, it allows to compare competing methods.

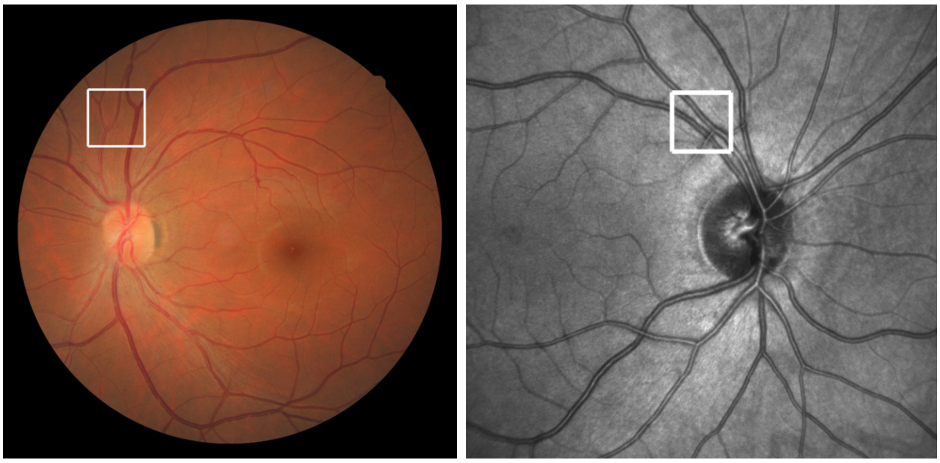


Figure 7.4: Reference images for $\mathcal{SR}1$ and $\mathcal{SR}2$. White square indicates detail shown in Figures 9.6 and 9.7.

7.4.1 3D registrations and shape estimation evaluation

To evaluate both 3D registration error, and the eye shape estimation error, the calculation presented in [133] is utilized. It measures the average distance between corresponding model points $\mathbf{x}_\kappa$ at the ground truth and the estimated solution. Given the ground truth $\mathbf{S}_g$ and an estimated solution $\mathbf{S}_h$, the model points for both are generated as described in Section 7.2.1, and the error is expressed as:

$$E = \frac{1}{\nu}\sum_{\kappa=1}^{\nu} |\mathbf{g}_\kappa - \mathbf{h}_\kappa|, \tag{7.1}$$

where $\mathbf{g}_\kappa$ are the 3D locations of the ground truth model points, $\mathbf{h}_\kappa$ the 3D locations of the estimated solution model points, and ν is the number of model points.

7.4.2 2D registration evaluation

The 2D evaluation of a registration method can be performed on two different levels. The first level is registration accuracy in a single image pair. The second is registration accuracy across a dataset.

Evaluation of a single image pair

We propose to calculate the registration error for an image pair as the mean distance between all $\mathbf{c}_j$ and $\mathbf{r}_j$ of the given image pair. The error E is the average of the distances d_j as formulated in Equation 7.2:

$$E = \frac{1}{n}\sum_{n} |\mathbf{c}_j - \mathbf{r}_j|, \tag{7.2}$$

where n is the amount of control points for the image pair.

Evaluation across a dataset

To evaluate the registration accuracy of a method across a dataset, we get inspired by relevant efforts in the field of evaluating object tracking methods [134]. More specifically, we utilize a 2D plot in the form of a Receiver Operating Characteristic (ROC) curve. The x axis of the plot corresponds to the value of an error threshold. If the registration error 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 creates a continuous, monotonic curve, which shows success rate as a function of target accuracy. Thus, the selection of an arbitrary threshold is avoided. Moreover, a registration method can 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.

By considering a variety of target registration accuracies, this measure facilitates the comparison of competing methods and the selection of the most appropriate one, given the registration accuracy requirements of a certain application. Such a plot is shown in Figure 8.8. When methods perform similarly, the Area Under Curve (AUC) of the plot as a percentage can be utilized. Table 8.6 shows the AUC for the results from Figure 8.8.

Making publicly available the registration error for each image pair is encouraged to facilitate the comparison between different registration methods. The registration errors for several methods are available at `http://www.ics.forth.gr/cvrl/fire`. The aim is to create a repository, where results obtained by other methods are included as they become available.

7.5 Discussion

RIR evaluation is a complex topic, given the lack of previously publicly available datasets that were suitable to this task, as well as the many different approaches to evaluate the registration error of a single image and across a dataset. Several contributions have been performed in this chapter.

First, a powerful yet simple tool has been presented. It allows not only to represent a retinal image in 3D if $\mathbf{S}$ is known, but also to generate a synthetic dataset with 2D images for which perfect ground truth is known. To the author's knowledge, no tool with these characteristics exists in the literature, nor is publicly available. The author has successfully utilized this tool to generate several synthetic datasets, from validating registration via relative pose $\{\mathbf{R}, \mathbf{t}\}$ estimation, to validating the eye shape estimation of all the utilized eye models.

Secondly, FIRE, a dataset with real images has been compiled and made publicly available (`http://www.ics.forth.gr/cvrl/fire`). FIRE consists of three categories of retinal image pairs. Each category is compiled with the intention of covering a different challenge in retinal image registration. For each image pair, ground truth information in the form of control points is provided.

Table 7.4 shows a comparison between e-ophtha, FIRE, RODREP and VARIA datasets. While e-ophtha, RODREP and VARIA present more standalone images and registrable image pairs, FIRE provides a wider range of scenarios, images with higher resolution and, most importantly, ground truth information, which is of paramount importance for the quantitative and, thus, objective assessment of registration accuracy.

Dataset	Images	FOV	Resolution	Registrable image pairs
e-ophtha [116], [117]	463	$\approx 45°$	2544×1696	144
FIRE [129]	129	45°	2912×2912	134
RODREP [130], [131]	1120	45°	2000×1312	≈ 1400
VARIA [126], [127]	233	20°	768×584	154
	Large overlap	Small overlap	Anatomical differences	Registration ground truth
e-ophtha [116], [117]	Yes	Yes	No	No
FIRE [129]	Yes	Yes	Yes	Yes
RODREP [130], [131]	Yes	Yes	No	No
VARIA [126], [127]	Yes	No	No	No

Table 7.4: Characteristics of image pairs in e-ophtha, FIRE, RODREP and VARIA.

We believe that the introduction of the FIRE dataset is a valuable contribution to the RIR community, as it provides data to perform fair comparisons, between retinal registration methods with different end applications in mind. Moreover, the suggested evaluation measure may assist researchers and/or practitioners in need of a registration method, to select the one that best suits their needs.

Chapter 8

REMPE configuration experiments

The main goal of the conducted experiments was to investigate how to configure all the elements from the Registration through Eye Modelling and Pose Estimation (REMPE) framework to form an accurate and robust Retinal Image Registration (RIR) method. Such experiments include investigating the appropriate keypoints, eye model, search space structure variant and Particle Swarm Optimization (PSO) budget distribution. A second goal was to perform a quantitative and comparative evaluation of the proposed approach to state of the art methods in RIR.

8.1 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 proposed RIR framework was implemented in C++ with openCV and CUDA tools.

The best trade-off between computational cost and accuracy for the framework is found to be when utilizing the following configuration: Information is retrieved from the image pairs using a combination of Scale-Invariant Feature Transform (SIFT) and vessel bifurcations. Solution **S** is calculated with $s = 3$ independent swarms consisting on a Random Sample Consensus (RANSAC) initialization using the spherical eye model, followed by a fine PSO search utilizing the ellipsoidal eye model. The fine PSO search is performed on a cube with the dimensions indicated in Equation 6.5. For each PSO process, a budget p / g = 10k / 300, is used. This configuration is utilized throughout the experiments, unless otherwise noted. A diagram of the method is shown in Figure 8.1.

Experiments are run on the Fundus Image Registration (FIRE) dataset, described in Section 7.3.1.

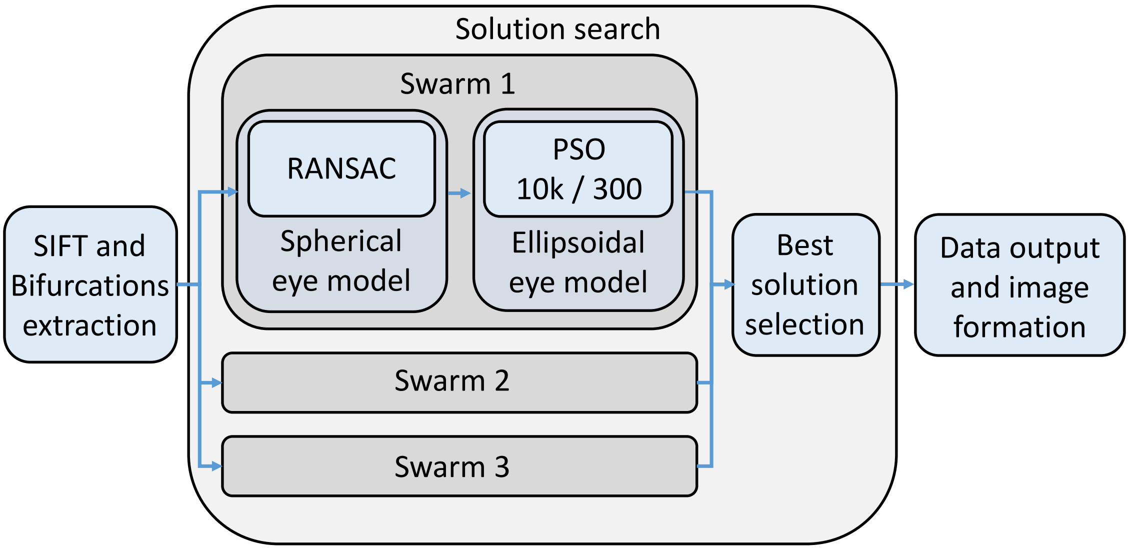


Figure 8.1: Workflow of the proposed registration method.

8.2 Keypoint selection

This experiment aims to compare the impact of the keypoints described in Section 5.1 on the accuracy of the proposed RIR framework. Table 8.1 shows the Area Under Curve (AUC) for every feature combination when registering the FIRE dataset [128]. Figure 8.2 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 Partial Intensity Invariant Feature Descriptor (PIIFD) on Harris corners, as in this case the framework provides results one step below the rest. A weak performance in 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

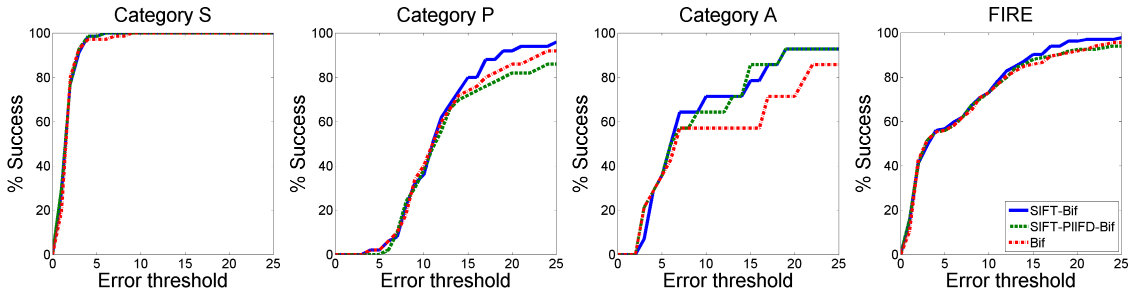


Figure 8.2: Registration success for SIFT-Bifurcations, SIFT-PIIFD-Bifurcations and Bifurcations. Results are presented as a Receiver Operating Characteristic (ROC) curve. The x axis marks the registration error threshold under which a registration is considered to be successful. The y axis marks the percentage of successfully registered image pairs for a given threshold.

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.

The results in this section, show 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. Speeded Up Robust Features (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. Given these results, the combination of SIFT and Bifurcations is chosen to extract corresponding keypoints from the image pairs.

8.3 Initialization

The aim of this experiment is to compare the performance between RANSAC [85] and posest [86] as initialization for the registration method. For both methods, the spherical eye model is used, as 2D-3D correspondences are required, and the spherical model is the most complex model for which $\{\mathbf{A}, \mathbf{Q}\}$ is known.

Table 8.2 and Figure 8.3 show that while the performance offered by both methods is similar, with the largest difference being in $\mathcal{A}$, generally RANSAC outperforms posest. As such, RANSAC was selected for initialization in our method.

SIFT	SURF	PIIFD	Bifurcations	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	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	**0.541**	**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 8.1: Percentage of AUC in each FIRE category for every keypoint combination. × indicates a keypoints type used for registration. Bold shows the highest score.

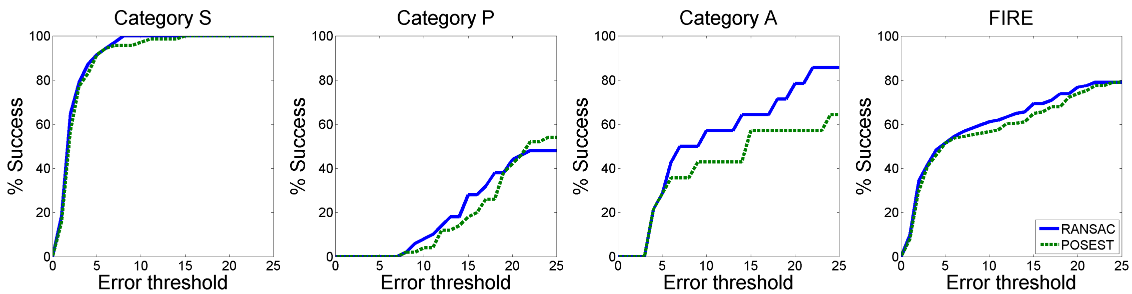


Figure 8.3: Registration success for RANSAC and posest with a spherical model. Figure interpretation in Figure 8.2.

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

Table 8.2: Percentage of AUC from Figure 8.4 in each FIRE category for RANSAC and posest with a spherical model.

8.4 Suitability of PSO

This experiment aims to prove the suitability of utilizing PSO over conventional projective geometry methods. The first method is affine transformation solved using linear algebra. The second method is a homography estimated using RANSAC. Given that the models utilized by those registration methods can be considered planar, the planar eye model is utilized in the proposed framework.

Table 8.3 and Figure 8.4 show that the proposed framework provides more accu-

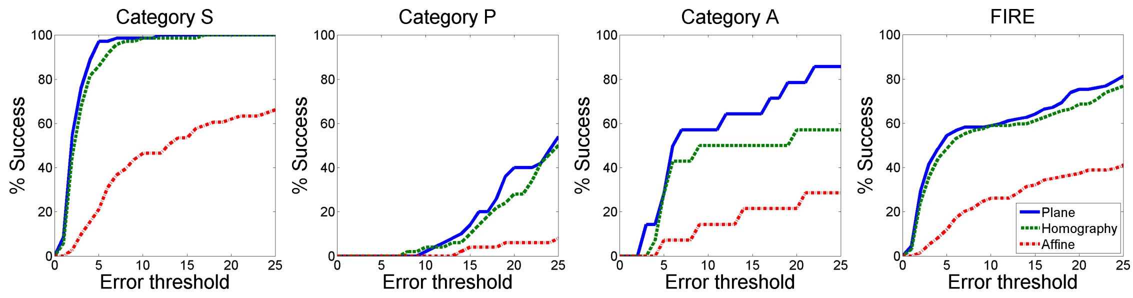


Figure 8.4: Registration success for homography and for PSO with a plane model. Figure interpretation in Figure 8.2.

rate registration results than the alternatives. Additionally, the proposed framework allows for more complex eye models, though presenting a higher potential for performing even more accurate registrations.

Method	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	FIRE
Plane	**0.926**	**0.164**	**0.574**	**0.607**
Homography	0.903	0.138	0.423	0.571
Affine	0.447	0.025	0.157	0.263

Table 8.3: Percentage of AUC from Figure 8.4 in each FIRE category for homography and for PSO with a plane model.

8.5 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 6.5 are compared. Variants C, R-C and R-F run p / g of 10k / 300 in a single PSO stage. Variants CF and R-CF run 10k / 150 in each of their two PSO stages. Thus, all variants, except R, have the same budget.

Table 8.4 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 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 $\mathcal{A}$. This experiment shows that with a good initialization, focusing the budget around the initial estimation provides the most accurate solution.

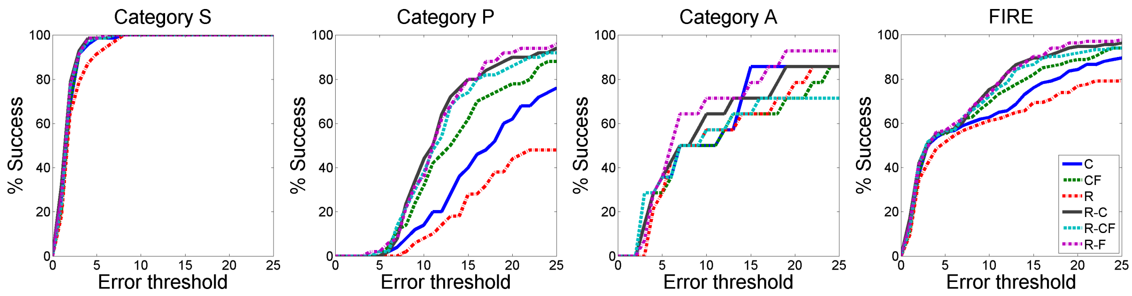


Figure 8.5: Registration success for for the 6 search space variants. Figure interpretation in Figure 8.2.

Variant	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	FIRE
C	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 8.4: Percentage of AUC from Figure 8.5 in each FIRE category for the 6 searchspace variants. Bold shows the highest score.

8.6 Impact of PSO budget

The aim of this experiment is to investigate the impact of key PSO parameters on the accuracy and performance of the registration method. A certain computational budget can be achieved with different combinations of p / g. However, the accuracy of the solutions reached by these combinations 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 Figure 8.6. Note that the x axis is logarithmic. For a given budget, distributing it across 300 generations provides the most accurate results, and that from a budget of 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 10k / 300 appears to be the best choice. The chosen budget is high compared to budgets used in other domains [89], [92], [93]. 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.

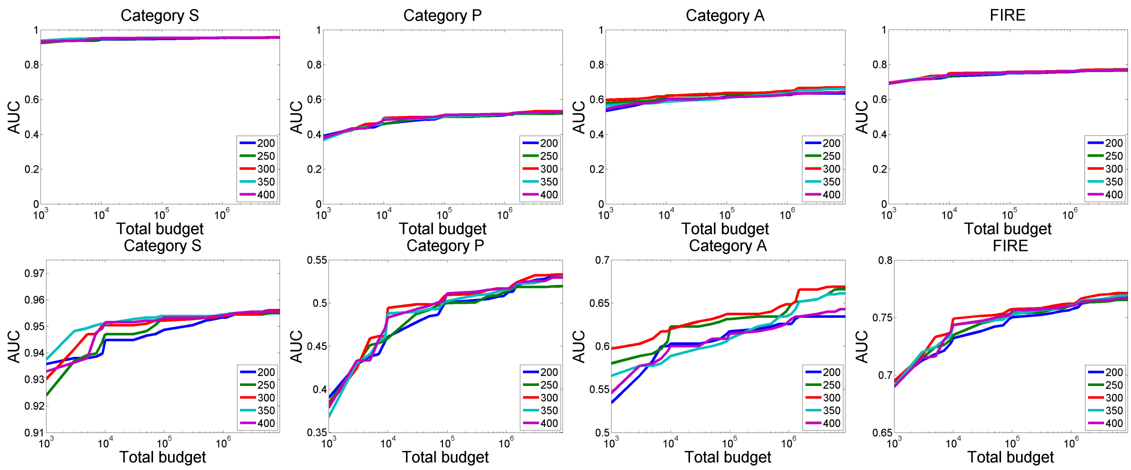


Figure 8.6: Budget study on registration success. y axis represents AUC for a given budget. x axis indicates the total budget used in the registration, in logarithmic scale.

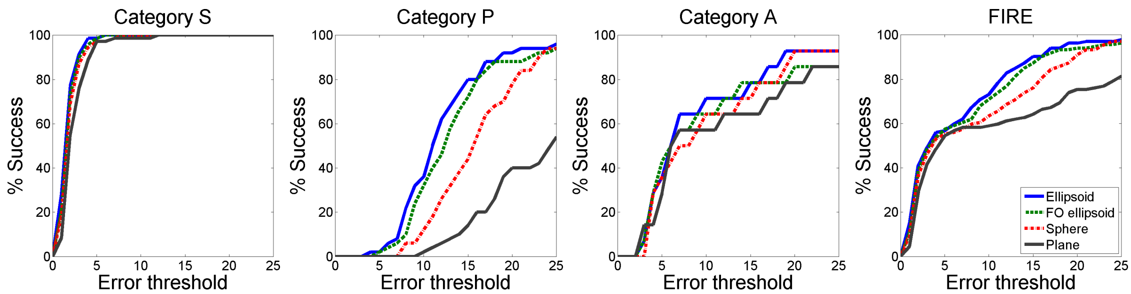


Figure 8.7: Registration success for for the four eye model variants. Figure interpretation in Figure 8.2.

8.7 Eye model variant comparison

This experiment compares the four variants proposed in Section 6.6 so as to identify which performs best in terms of registration accuracy. Results are shown in Table 8.5 and Figure 8.7.

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 datasets 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.

8.8 Multiple swarms

This experiment evaluates the effect of executing the RANSAC and PSO process several parallel times, to palliate the non-deterministic effect of the registration

Model	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	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**

Table 8.5: Percentage of AUC from Figure 8.7 in each FIRE category for the four eye model variants. Bold shows the highest score.

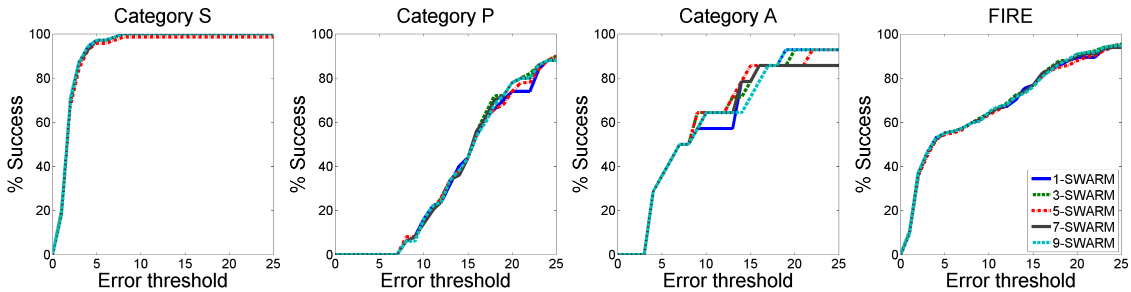


Figure 8.8: Registration success for an increasing amount of swarms. Figure interpretation in Figure 8.2.

framework. The framework is executed 10 times, with each execution having a different s ranging from 1 to 10 swarms.

Table 8.6 shows the AUC for every execution when registering the FIRE dataset. Figure 8.8 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, $s = 3$ is selected, as it provides the best robustness computational performance trade-off.

Swarms	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	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 8.6: Percentage of Area Under the Curve (AUC) in Figure 8.8 in each FIRE category for an increasing number of independent swarms. Bold shows the highest score.

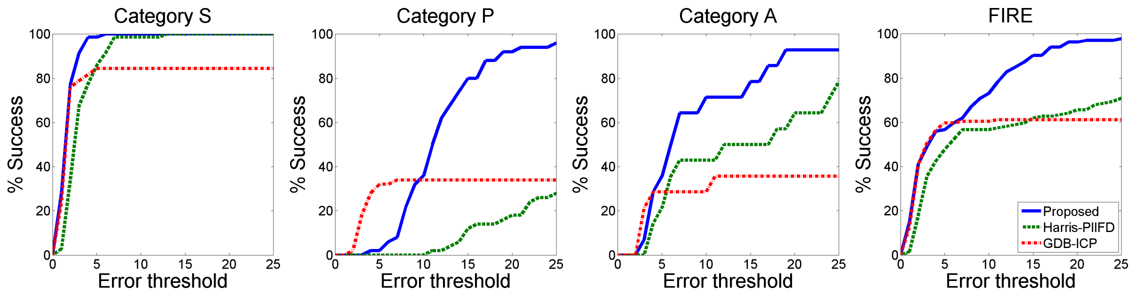


Figure 8.9: Registration success for the proposed method, Harris-PIIFD [37] and GDB-ICP [49]. Figure interpretation in Figure 8.2.

8.9 Comparison with state of the art

In this experiment, the accuracy of the registration method is compared to two state-of-the-art methods. The methods are Generalized Dual Bootstrap-Iterative Closest Point (GDB-ICP) [49] and the original Harris-PIIFD framework [37]. These methods have been selected as both are widely applied in the retinal image registration field. GDB-ICP is used in [36], [37], [45], [53] and the original Harris-PIIFD framework in [29], [55]. For GDB-ICP, the C++ implementation provided by the authors[1] was used. For the original Harris-PIIFD framework, a MATLAB implementation was utilized.

Figure 8.9 and Table 8.7 show the result plot and the AUC, respectively. The proposed framework outperforms the original Harris-PIIFD framework [37] in every category. It outperforms GDB-ICP [49] except up to a certain threshold in $\mathcal{P}$.

Method	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$	FIRE
Proposed	**0.958**	**0.542**	**0.660**	**0.773**
Harris-PIIFD [37]	0.900	0.090	0.443	0.553
GDB-ICP [49]	0.814	0.303	0.303	0.576

Table 8.7: Percentage of AUC from Figure 8.9 in each FIRE category for the proposed method, Harris-PIIFD [37] and GDB-ICP [49]. Bold shows the highest score.

8.10 Execution time

This experiment aims to perform a comparison regarding execution times of the proposed method, as well as GDB-ICP [49] and the original Harris-PIIFD framework [37]. It also studies the execution times of diverse parts of the proposed algorithm, to identify potential bottlenecks.

Table 8.8 shows average registration time per image pair, in seconds, for the proposed method, Harris-PIIFD and GDB-ICP. While Harris-PIIFD is the fastest

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

performing method of the three, in Section 8.9 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, but it is also shown to provide the most accurate registration of the three methods. RIR is a medical application, and as such, emphasis is placed on accuracy rather than computational time. Besides that, while REMPE takes longer than competing methods to perform registration, it is still within acceptable limits, as it allows to register image pairs within an examination session.

	Proposed	Harris-PIIFD [37]	GDB-ICP [49]
Time (s)	198	11	73

Table 8.8: Average registration time per image pair, in seconds, for the proposed method, Harris-PIIFD [37] and GDB-ICP [49].

Table 8.9 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. Then the PSO process, which has been implemented utilizing GPU acceleration. There exist apparent bottlenecks in the keypoint extraction and matching, and the 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.

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

Table 8.9: Average process time for diverse stages of the proposed method.

While the proposed method is slower than competing methods, it provides more accurate registration and the execution time is within reasonable limits. Bottlenecks on the bifurcation extraction, as well as the image formation, have been identified, and will be addressed in future work.

8.11 Multiple image registration

This experiment aims to qualitatively show that the proposed RIR framework allows for the simultaneous registration of multiple F_{t_i} to a single F_0. While this can be performed in 2D with multiple pairwise registrations of F_{t_i} to F_0, this task is not trivial for the case of eye shape estimation. The simultaneous registration presented

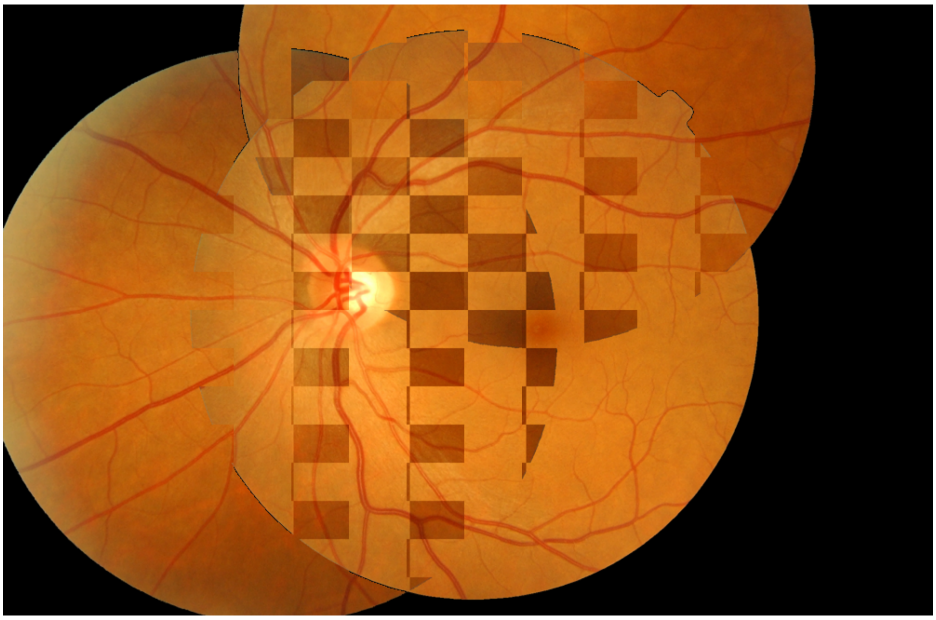


Figure 8.10: 2D Simultaneous registration of multiple images.

here demonstrates that the proposed framework allows for the combination of information from multiple images to perform a single $\{\mathbf{A}, \mathbf{Q}\}$ estimation. Thus, it allows to estimate a single $\mathcal{E}$ and register all images to it in 3D.

Experiments have been performed utilizing images from the RODREP [130] dataset. This is due to the fact that it has been compiled with the purpose of performing registration of multiple images from the same eye for creating mosaics, thus making it the most suitable dataset for this purpose. Figures 8.10 and 8.11 show qualitative results for this type of multiple registration.

8.12 Discussion

The performed experimental evaluation validates the proposed method, publicly available at `http://www.ics.forth.gr/cvrl/rempe`. The combination of SIFT and bifurcations is shown to allow for more accurate registration than standalone features. For planar models, PSO is shown to provide improved results over alternatives. Regarding eye models, an ellipsoid, which has the potential to approximate the actual shape of the eye more accurately than alternative models such as an sphere or a plane is also shown to improve accuracy. Finally, experiments show the proposed method to be more accurate and robust than state-of-the-art methods such as GDB-ICP and Harris-PIIFD.

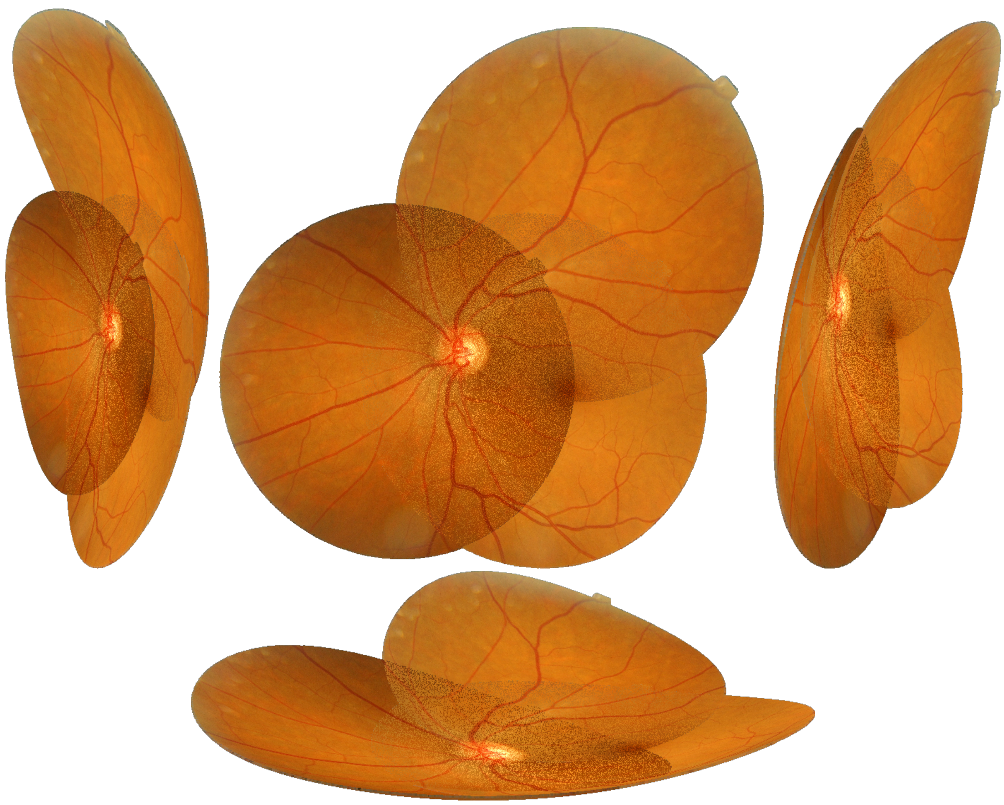


Figure 8.11: 3D Simultaneous registration of multiple images.

Chapter 9

REMPE registration applications

In this chapter, the suitability of utilizing Registration through Eye Modelling and Pose Estimation (REMPE) in four applications of Retinal Image Registration (RIR) has been explored. In particular, for the cases of longitudinal studies, mosaicing, Super Resolution (SR) and eye shape estimation. The reason behind selecting these four applications is the suitability of the results to be directly utilized by clinicians in their work.

9.1 Longitudinal studies

These experiments evaluate the applicability of the proposed method in longitudinal studies. Figure 9.1 shows registration results for 4 image pairs of the Fundus Image Registration (FIRE) [129] dataset. Image pairs from the top row correspond to category $\mathcal{S}$, so they contain no anatomical differences, and the pairs from the bottom row belong to category $\mathcal{A}$, so there exist anatomical differences within each image pair.

These results show that the proposed RIR method performs accurate registration for image pairs in which there may exist 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.

9.2 Mosaicing

These experiments show mosaicing results for the proposed registration method for 11 image pairs from the 4 publicly available datasets that present registrable image pairs, as discussed in Section 7.1 and whose characteristics are described Table 7.4. Figure 9.2 shows results for 2 images pairs from category $\mathcal{P}$ of the FIRE [129] dataset. Figure 9.3 displays results for 3 image pairs from the RODREP [131]

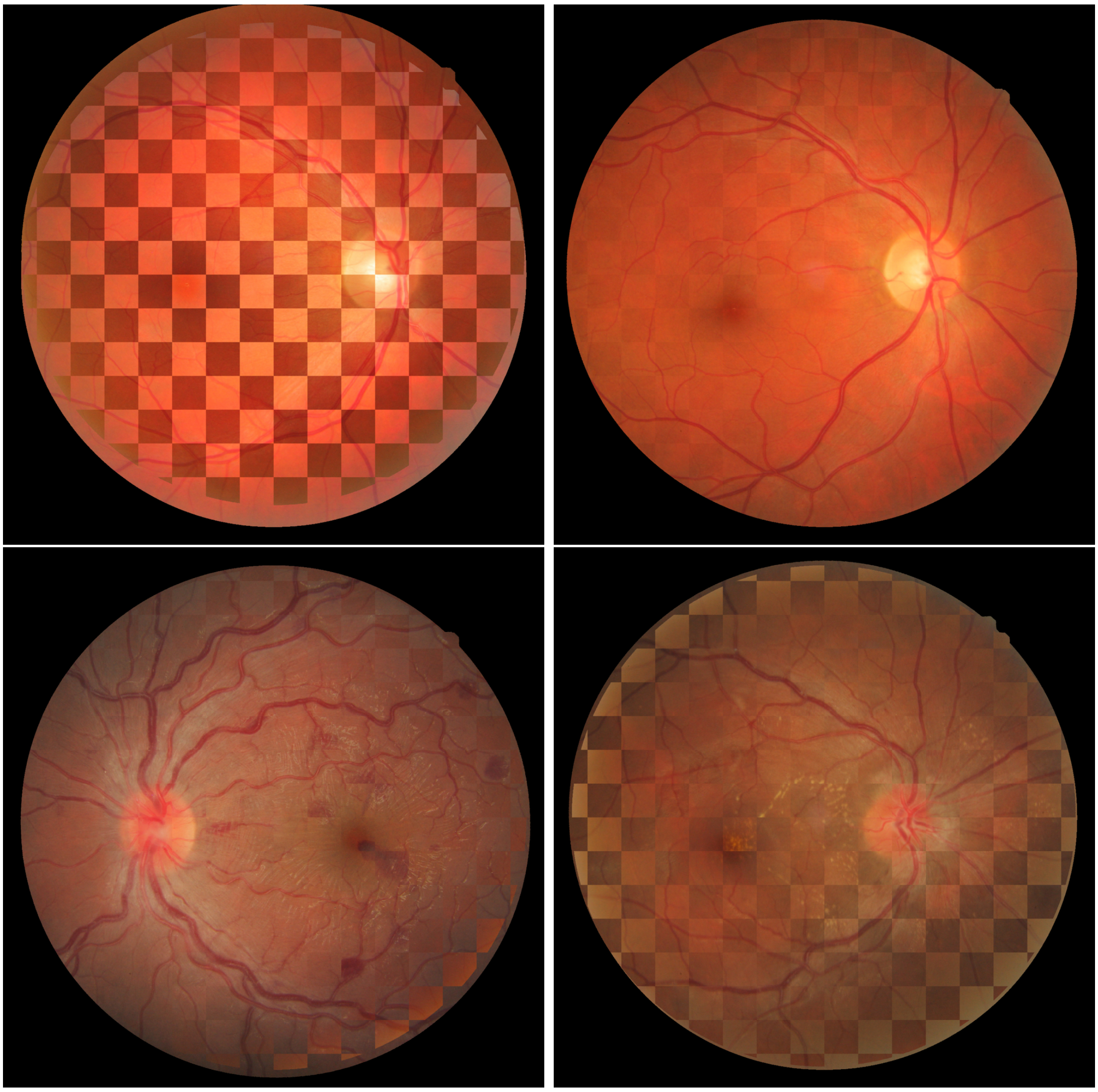


Figure 9.1: Longitudinal registration results for the FIRE dataset. Top row shows registration for image pairs from category $\mathcal{S}$, with no anatomical differences, while bottom row shows registration for category $\mathcal{A}$, with anatomical differences. Results displayed in a checkerboard mosaic, alternating patches from both images, with the intent to show registration accuracy, not for analyzing anatomical differences in the registered images.

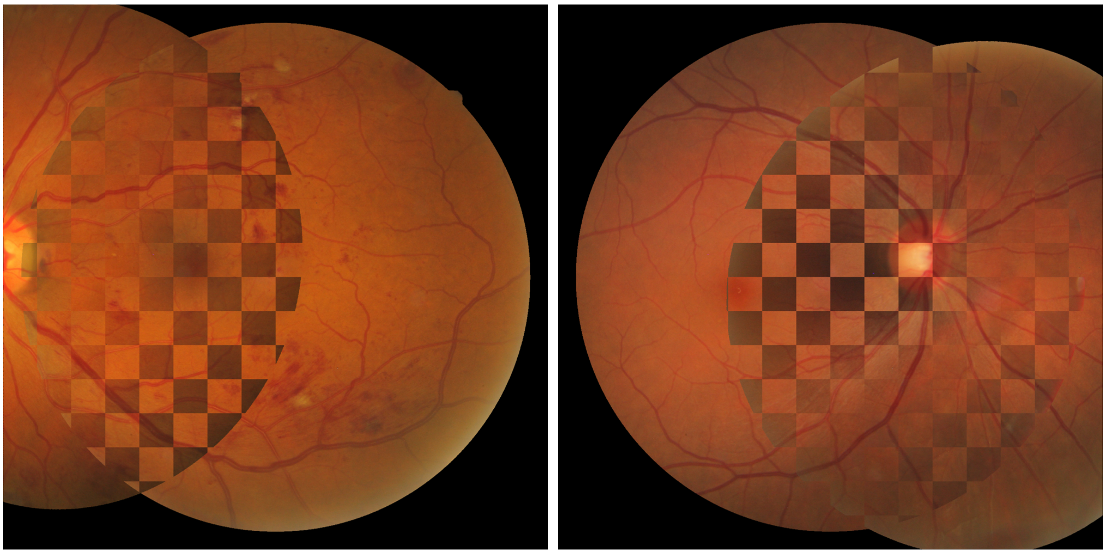


Figure 9.2: Mosaic registration results for category $\mathcal{P}$ of the FIRE dataset. Figure interpretation in Figure 9.1.

dataset. Figure 9.4 presents results for 3 image pairs from the e-ophtha [117] dataset. Finally, Figure 9.5 shows results for 3 image pairs from the VARIA [127] dataset.

These results show that the proposed RIR framework performs accurate registration for image pairs with a wide range of characteristics. Image pairs may present varying degrees of overlap, Field of View (FOV), resolution and they may even be monochromatic. As discussed in Section 4.2, mosaicing is an important tool for clinicians, so providing accurate mosaics is of utmost importance for a RIR framework.

9.3 Super resolution

In this work, while we do not focus on the suitability of multi-frame SR for obtaining images of higher resolution and definition, we demonstrate that the proposed RIR method leads to improved SR images when compared to state-of-the-art RIR methods. We have comparatively evaluated the proposed method, as well as Generalized Dual Bootstrap-Iterative Closest Point (GDB-ICP) [49] and Partial Intensity Invariant Feature Descriptor (PIIFD) on Harris corners [37] as base registration methods. For SR, the MATLAB implementations[1] for the methods presented in [66], [135] and [136] as well as interpolation of image points were utilized.

Multi-frame SR is quantitatively evaluated in relevant literature with a range of metrics. In [70] both Structural Similarity (SSIM) and Mean Square Error (MSE) are utilized. Peak Signal-to-Noise Ratio (PSNR) is used in [137]. Signal-to-Noise Ratio (SNR) is applied in [6]. As these methods require to compare with a reference image, the typical procedure consists to reduce the size of the images to register,

[1] `http://lcav.epfl.ch/software/superresolution`

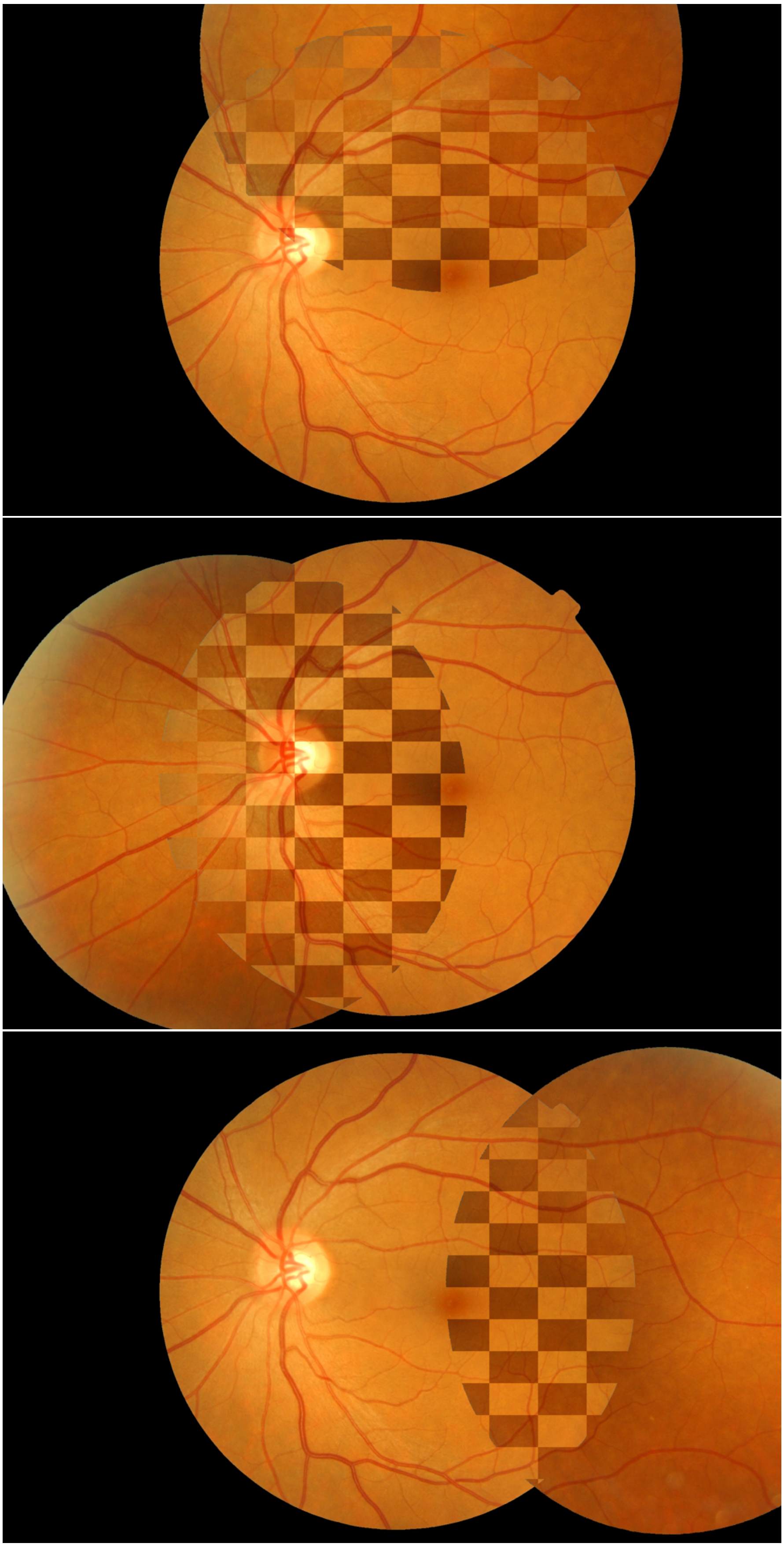


Figure 9.3: Registration results for the RODREP dataset. Figure interpretation in Figure 9.1.

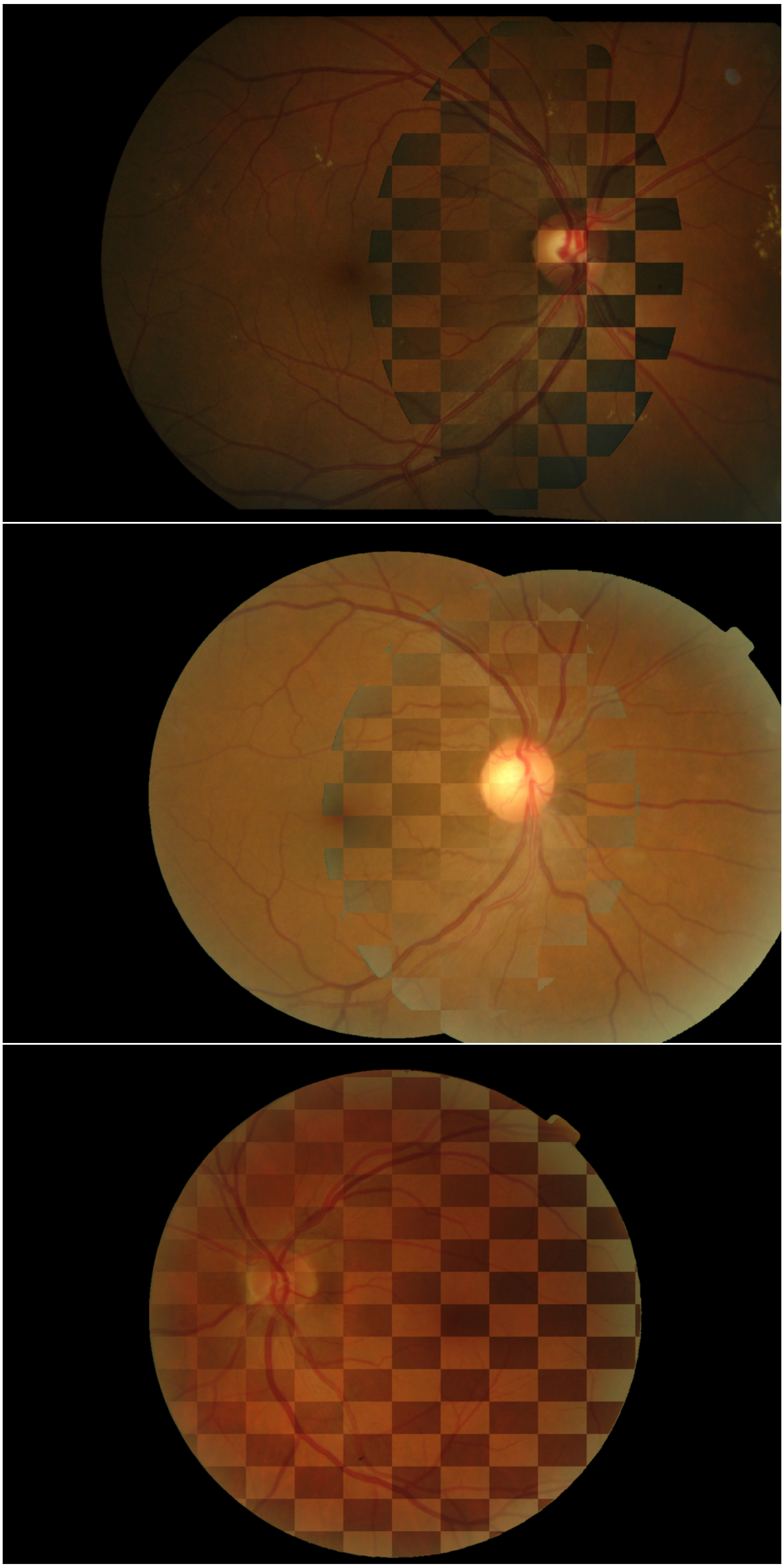


Figure 9.4: Registration results for the e-ophtha dataset. Figure interpretation in Figure 9.1.

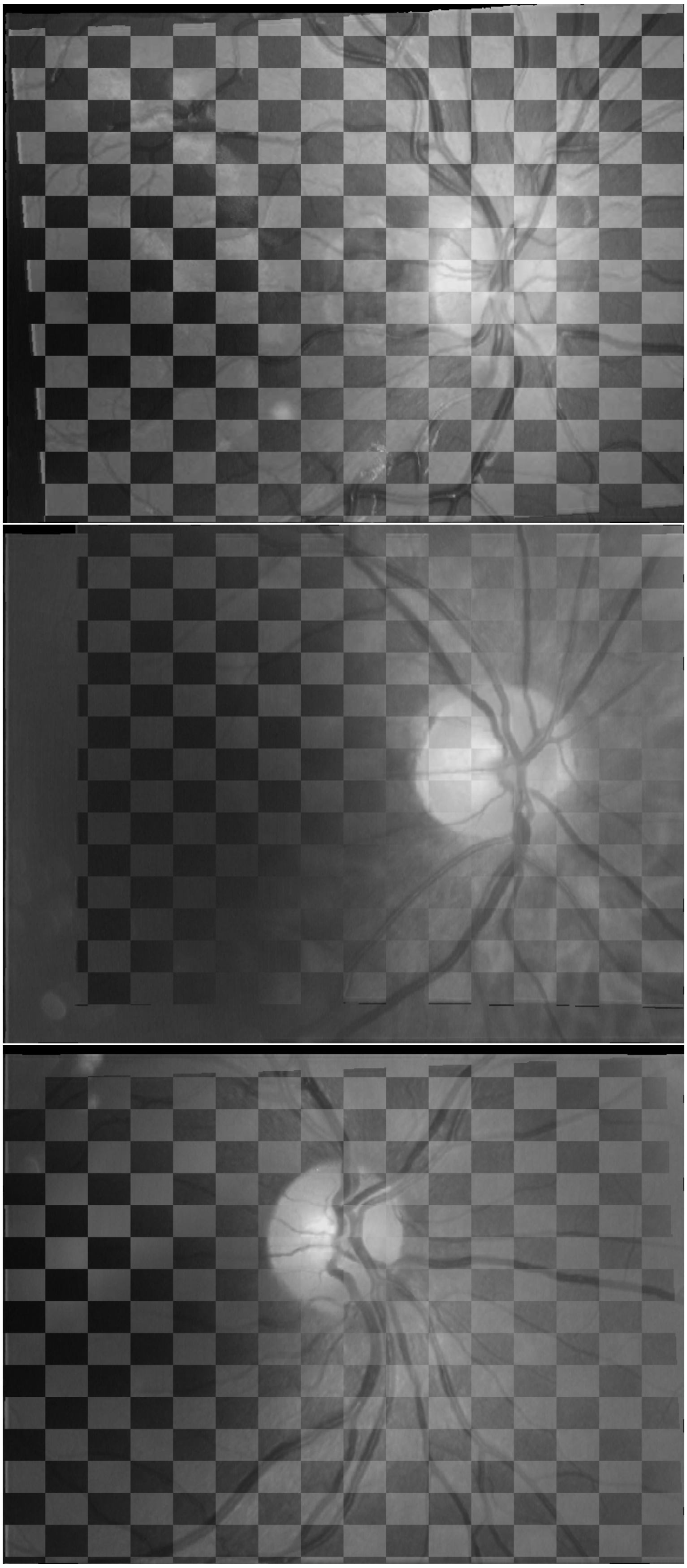


Figure 9.5: Registration results for the VARIA dataset. Figure interpretation in Figure 9.1.

so that the resulting SR image is of equal size to the original reference image. In this case, the images were resized to one third of their original size and SR was performed with scaling factor $n = 3$. $l = 9$ images were used in the registration and SR process.

Results for dataset $\mathcal{SR}1$ are shown in Tables 9.1, 9.2, 9.3 and 9.4. Results for dataset $\mathcal{SR}2$ are shown in Tables 9.5, 9.6, 9.7 and 9.8. These tables show that for a given SR method, results are best when utilizing the proposed RIR method. This stems from the fact that the proposed method performs registration more accurately than the competing methods.

Method	[66]	[135]	[136]	Interpolation
Proposed	**56.568**	**56.515**	**55.973**	**56.515**
GDB-ICP [49]	55.420	55.257	54.507	55.256
Harris-PIIFD [37]	39.385	39.308	33.827	39.307

Table 9.1: SNR for SR comparison in $\mathcal{SR}1$. Bold shows the best score.

Method	[66]	[135]	[136]	Interpolation
Proposed	**2.960**	**2.942**	**2.373**	**2.942**
GDB-ICP [49]	2.526	2.454	1.792	2.453
Harris-PIIFD [37]	-4.862	-4.893	-7.003	-4.894

Table 9.2: PSNR for SR comparison in $\mathcal{SR}1$. Bold shows the best score.

Method	[66]	[135]	[136]	Interpolation
Proposed	**0.932**	**0.930**	**0.930**	**0.930**
GDB-ICP [49]	0.930	0.929	0.929	0.929
Harris-PIIFD [37]	0.881	0.879	0.859	0.879

Table 9.3: SSIM for SR comparison in $\mathcal{SR}1$. Bold shows the best score.

Figures 9.6 and 9.7 respectively show detail of the results for $\mathcal{SR}1$ and $\mathcal{SR}2$ utilizing the [66] SR method. From left to right: reference image, proposed method, GDB-ICP [49] and Harris-PIIFD [37].

9.4 Eye shape estimation

These experiments were performed to validate the eye shape estimated by the registration framework. Modalities such as Computed Tomography or Magnetic Resonance could provide ground truth for the actual shape of a given eye. However such data is unavailable to us. Instead, synthetic images with known ground truth $\mathbf{S}$ were generated.

Method	[66]	[135]	[136]	Interpolation
Proposed	**189.658**	**190.009**	**201.124**	**190.004**
GDB-ICP [49]	198.072	199.516	213.162	199.528
Harris-PIIFD [37]	414.664	415.958	513.653	415.985

Table 9.4: MSE for SR comparison in $\mathcal{SR}1$. Bold shows the best score.

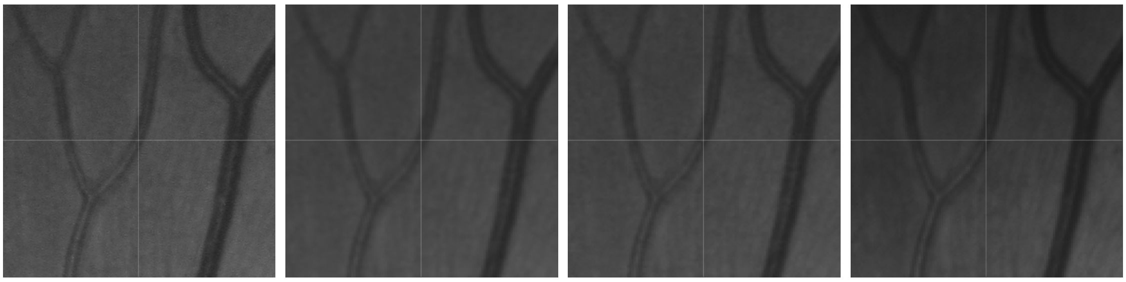


Figure 9.6: Detail of SR results for $\mathcal{SR}1$. From left to right: reference image, proposed method, GDB-ICP [49] and Harris-PIIFD [37].

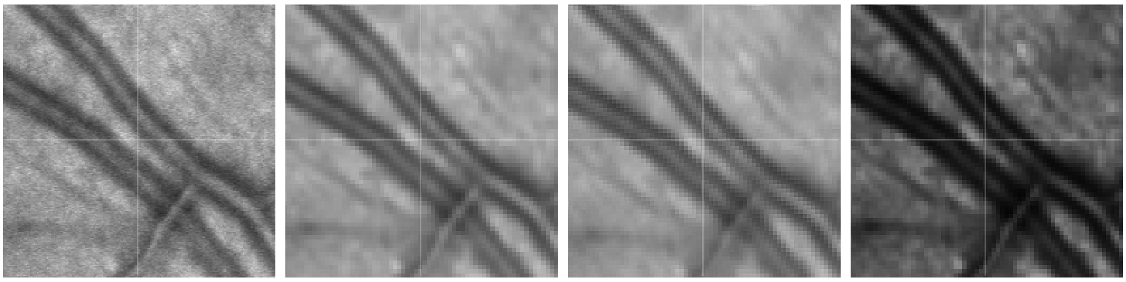


Figure 9.7: Detail of SR results for $\mathcal{SR}2$. From left to right: reference image, proposed method, GDB-ICP [49] and Harris-PIIFD [37].

Two sets of synthetic images were generated utilizing a fixed orientation ellipsoid model. The images were generated utilizing the same $\mathbf{S}_i$, but with different virtual cameras, with FOV of 45° and 100° respectively. The proposed framework is utilized to estimate $\{\mathbf{R}, \mathbf{t}\}$ and $\mathbf{A}$. For both datasets, the framework is run twice. In one run, all 9 parameters are estimated. In the other, one of the axes of the eye model is fixed to the ground truth, so only 8 parameters are estimated. As the fixed orientation ellipsoid model is used, the model rotation parameters are fixed to 0 and not estimated.

Results are shown in Table 9.9. When one of the axes of the model is known, the framework is able to correctly estimate $\mathbf{S}$ with high accuracy for both FOV. However, when the three axes from $\mathbf{A}$ are unknown, the framework is not able to estimate $\mathbf{A}$ for a FOV of 45°. These results prove that with a wide field of view, the framework is able to perform 3D registration with an accurate estimation of the eye model. If the Navarro eye model is utilized, that means an average error of 0.06 mm. However, for 45°, which is the FOV of interest in this work, the error grows to 0.78 mm.

Method	[66]	[135]	[136]	Interpolation
Proposed	**100.299**	**98.622**	**99.260**	**101.252**
GDB-ICP [49]	83.167	84.056	78.988	84.892
Harris-PIIFD [37]	-8.240	-8.423	-8.261	-8.394

Table 9.5: SNR for SR comparison in $\mathcal{SR}2$. Bold shows the best score.

Method	[66]	[135]	[136]	Interpolation
Proposed	**10.684**	**9.876**	**10.164**	**11.177**
GDB-ICP [49]	2.247	2.710	0.122	3.120
Harris-PIIFD [37]	-28.285	-28.314	-28.307	-28.303

Table 9.6: PSNR for SR comparison in $\mathcal{SR}2$. Bold shows the best score.

After these results, a second experiment was conducted on the 45° images to analyze if the erroneous shape estimation was due to keypoint localization error. In this case, the keypoints from the framework were substituted by perfect correspondences. Each pixel in F_0 is a keypoint, and as the ground truth is known, it's correctly matched to its location in F_t. Additionally, this same experiment is conducted adding a localization errors of $\epsilon < 0.25$, $\epsilon < 0.5$ and $\epsilon < 1$ pixel to each keypoint. The experiment consisted on comparing the score of the objective function of the $\mathbf{S}_h$ found by the registration framework with the score of the ground truth.

For $\epsilon = 0$, the ground truth location presented the optimal score, with $\mathbf{S}_h$ approximating the ground truth values for $\{\mathbf{R}, \mathbf{t}\}$ and $\mathbf{A}$. However, for $\epsilon < 0.25$, $\epsilon < 0.5$ and $\epsilon < 1$, the ground truth location does not present the optimal score, and while $\mathbf{S}_h$ approximates the ground truth values for $\{\mathbf{R}, \mathbf{t}\}$, it deviates for $\mathbf{A}$.

This shape estimation error is seen in synthetic images that were generated using an eye model with a smooth surface. In real images, the retinal surface is irregular. The impact of this surface irregularity in the 3D keypoint localization and shape estimation leads us to believe that the error in the shape estimation will be larger in real images than in synthetic ones. Given the need of clinicians for accurate measurements, we believe that at this stage of the research the performance of measurements using real distances (i.e. mm) in the 3D estimated models is not yet accurate enough. This eye shape estimation error does not affect the accuracy for 2D registration and measurements in pixels in 2D images, as shown in previous experiments.

9.5 Discussion

In this chapter, the suitability of REMPE for RIR applications such as longitudinal studies, mosaicing and SR is demonstrated. For longitudinal studies, qualitative results are shown both for image pairs with and without anatomic differences. For

Method	[66]	[135]	[136]	Interpolation
Proposed	**0.956**	**0.960**	**0.956**	**0.969**
GDB-ICP [49]	0.795	0.817	0.784	0.816
Harris-PIIFD [37]	0.636	0.626	0.634	0.626

Table 9.7: SSIM for SR comparison in $\mathcal{SR}2$. Bold shows the best score.

Method	[66]	[135]	[136]	Interpolation
Proposed	**87.606**	**94.978**	**92.285**	**83.393**
GDB-ICP [49]	203.691	194.471	251.922	186.658
Harris-PIIFD [37]	4314.5	4327.1	4324.2	4322.3

Table 9.8: MSE for SR comparison in $\mathcal{SR}2$. Bold shows the best score.

	45°	100°
8 DoF search	0.25%	0.06%
9 DoF search	6.54%	0.52%

Table 9.9: Error on the estimation of $\mathbf{A}$ in synthetic images. Error is indicated as the average of the percentage of the ground truth values.

mosaicing, qualitative results are shown for the proposed method when registering images across 4 different publicly available datasets. Additionally, REMPE is quantitatively shown to outperform state-of-the-art methods when utilized to register images to be used with multi-frame SR methods. Finally, a preliminary study on the general eye shape estimation is presented. While in this last application the performance is not yet as accurate as it would be desired, it is a hopeful start in an application that we believe can improve the measurements in which clinicians base their diagnostics.

Part III

Future work and discussion

Chapter 10

Future work

In this chapter, limitations of the current approach are described and lines of work to address them are proposed.

10.1 Improvement of common information retrieval

Typically, retinal images have very little texture, and the highest textured areas are located around the blood vessels and the optic disc. Highly textured areas are also present as some forms of retinopathy, such as hemorrhages or cotton-wool spots. However, these forms of retinopathy are not present in every eye, and their shape and size may vary between examinations, thus they are not fully reliable visual features. These facts lead to certain areas of the image pairs containing high density of correspondences, while others have low density. In turn, this indicates that certain areas of the image have more weight in the final registration process. Table 10.1 presents some of the challenges present on the Fundus Image Registration (FIRE) dataset. A number of approaches to solve these issues exist, such as analyzing the localization of correspondences, utilizing a progressive ratio for matching or exploring global cues.

Correspondences	$\mathcal{S}$	$\mathcal{P}$	$\mathcal{A}$
Low amount		$\times\times$	$\times\times\times$
Uneven distribution	$\times$	$\times$	$\times\times\times$
Erroneous matches	$\times$	$\times$	$\times\times\times$

Table 10.1: Difficulties of keypoint location and matching in the three categories of the FIRE dataset.

10.1.1 Density of correspondence location

Equation 6.2 shows that each corresponding keypoint pair weights the same when calculating the score of the objective function. However, keypoint correspondences can be heavily clustered in certain areas of the image pairs. This means that the registration is biased towards these heavily clustered areas, potentially leading to a less accurate registration overall. This issue can be approached in several ways. The first alternative would be to remove matches that are closely located to each other. This would work towards evening out the keypoint distribution across the image. The second one would be to introduce a weight for each of the corresponding pairs. Said weight would be inversely proportional to the amount of closely co-located matches. Alternatively, Adaptive Non-Maximal Suppression (ANMS) [138] or Suppression via Disk Covering (SDC) [139], methods developed for better distributing the keypoints over the whole image, avoiding clustered matches, could be utilized.

10.1.2 Progressive ratio for matching

As discussed in Section 5.1.5, bilateral matching based on the conventional method introduced by Lowe in [81] is utilized. This technique uses a fixed distance ratio when finding correspondences in the keypoints. While this fixed ratio proves to be adequate for image pairs like the ones in $\mathcal{S}$, for some pairs in $\mathcal{P}$ and $\mathcal{A}$ this leads to a low number of correspondences. A potential way to overcome this issue, would be to gradually increase the ratio until a minimum amount of correspondences is reached.

10.1.3 Alternative distance in matching algorithm

Another alternative to increase the amount of correspondences, besides the progressive ratio for matching described in the previous subsection, consists on utilizing other distance functions for the matching process. Currently the conventional method by Lowe [81] with Euclidean distances is utilized. Alternative distance functions such as Chi-square have been successfully employed in pose estimation problems [86], and the proposed method could benefit from them.

10.1.4 Exploration of global cues

The primary reason to not explore the use of global cues is that in images with high resolution, their utilization is impractical within the proposed framework due to execution time. The development of parallel implementations of these cues and the utilization of GPUs would decrease the execution time of the registration process. Then, a more exhaustive analysis of the accuracy and robustness provided by these global cues either in isolation or combined among themselves or even with local methods could be performed. This has the potential to improve the registration results provided by the framework.

10.2 Improvement of camera and eye models

An important criticism that can be done to the proposed framework is that big simplifications are performed on the camera, eye shape and eye optic models. More accurate modeling of these elements has the potential to lead to better image registration.

10.2.1 Camera model

The camera model could be improved by performing auto-calibration. Auto-calibration is a process that allows to determine the internal camera parameters from multiple uncalibrated images from a scene. This is conventionally performed via the Kruppa equations. The applicability of this method had been explored with regard to retinal images in a few works.

Matinez-Perez and Espinosa-Romero [73] use the simplification via singular value decomposition (SVD) of the Kruppa equations [76] described by Lourakis and Deriche [140] to calculate the fundamental matrix.

Laliberté et al [71] compare several methods for the calculation of the fundamental matrix, but claim that the best result they obtain is using their implementation of the 8-point algorithm.

Choe et al [12] test the 7-point and 8-point algorithms. They obtain bad results because in narrow Field of View (FOV) fundus images, the imaged surface is relatively flat. Even though the points are not on the same plane, the 3D depth range is too narrow. They find success when utilizing the plane-and-parallax algorithm from Kumar et al [75].

Lin and Medioni [11], similarly to Choe [12], use the plane-and-parallax algorithm [75].

10.2.2 Eye optic model

The eye optic model utilized by our registration method is a very simple one, as it only considers the general shape of the retina, and ignores other elements such as the cornea, the lens and the aqueous humor. This simplification allows us to shoot straight rays from the camera pinhole, through the 2D retinal image and straight to the retina. However, the three mentioned elements comprise a complex optic system in which the light rays do not follow a straight path.

10.2.3 Eye shape model

The geometrical eye models utilized are simple geometrical models with a smooth surface, derived from an ellipsoid. However, the surface of the retina is not smooth, but has an irregular surface due to the presence of vessels and tissue irregularities. Additionally, the optic disc has a cupped shape. If these particularities are taken into account, a more accurate 3D model of the eye can be estimated, leading to improvements in the registration of retinal images.

10.3 Quaternions for rotation parameterization

Rotations are parameterized utilizing Euler angles. However, there is a number of shortcomings to this notation, such as gimbal lock. In a gimbal lock, the three gimbals reach a parallel configuration, locking the system into a degenerate two-dimensional space. In the proposed method we deal with small rotations; and the way that they parameterization is set, the gimbal lock is guaranteed not to show up due to anatomical constraints. However, utilizing quaternions for the rotation parameterization would completely eliminate this risk.

10.4 Parallelization

The Particle Swarm Optimization (PSO) process has been parallelized using GPU acceleration. However, as shown in Section 8.10, bottlenecks have been identified in the bifurcation extraction and image formation parts of the algorithm. These bottlenecks would benefit from parallelized implementations, leading to a decrease of the overall execution time of the algorithm.

10.5 Applications

Having an accurate registration method, an organic way to continue research in this field is to use it for the applications described in Section 4. Some application results have been presented in this work, but there are several additional opportunities.

10.6 Discussion

While the proposed Retinal Image Registration (RIR) framework is shown to outperform state-of-the-art methods, there is still quite some room for improvement. In this chapter, several improvements such as the retrieval of common information in the image pair, the keypoint matching method or the camera and eye models are analyzed, and possible approaches to approach them are described.

Chapter 11

Discussion

Retinal Image Registration (RIR) is a topic of high interest for clinicians and medical experts. This is both due to its direct application to the images they work with, as well as a base for applications such as image enhancement, mosaicing, and longitudinal studies.

In this thesis, the problem of RIR is described and prior work in the topic discussed. The Registration through Eye Modelling and Pose Estimation (REMPE) framework is proposed and has been made publicly available at `http://www.ics.forth.gr/cvrl/rempe`. It performs registration by simultaneously estimating the pose of the cameras utilized to acquire the retinal images, as well as the general shape and pose of the eye. This represents a novel approach as typically RIR methods warp the test image either considering the retina as a plane, or utilizing a quadratic function without regard for the retina's actual shape. The proposed approach is evaluated quantitatively and shown to outperform state-of-the-art methods.

The application of the proposed registration method to tasks of high clinical importance such as longitudinal studies, mosaicing and super resolution is also explored, obtaining satisfactory results. Additionally, the suitability of utilizing the proposed method for accurate eye shape estimation with the purpose of performing measurements in the eye model is probed, showing promising results while still having plenty of room to improve.

A set of tools for generating synthetic, realistic 3D eye models, is presented. These tools also allow to create 2D retinal images from different views. This contribution allows for the validation of the geometry of the system comprised by the eye model and the virtual cameras, as well as for evaluating registration accuracy.

The Fundus Image Registration (FIRE) dataset [128], comprised of pairs of real retinal images has been compiled and made publicly available at `http://www.ics.forth.gr/cvrl/fire`. FIRE consists of three types of images. Each category is compiled with the intention of covering a different challenge in retinal image registration. For each image pair, ground truth information in the form of control points is provided. This dataset is a notable contribution to the RIR community, given the previous lack of datasets providing ground truth for facilitating registration accuracy analysis and comparison.

11.1 Impact

The work presented in this thesis has been part of the Retinal Vascular Modeling, Measurement And Diagnosis (REVAMMAD) project. REVAMMAD is an interdisciplinary European Commission project (Project number 316990) with the aim of combating some of the European Union's most prevalent chronic medical conditions using retinal imaging. It involved researchers from 10 institutions in 7 European countries. This work has also been part of the FORTH-ICS internal RTD Programme "Ambient Intelligence and Smart Environments".

11.1.1 Thesis-related publications

Parts of the work presented in this thesis have led to the publication of articles in several peer-reviewed scientific journals and conferences, as well as the public release of a dataset and the software itself. To date[1], this work has been cited 16 times, and has an h-index of 3[2].

Software

- REMPE: Registration through Eye Modelling and Pose Estimation.
  Available at `http://www.ics.forth.gr/cvrl/rempe`.

Dataset

- FIRE: Fundus Image Registration dataset.
  Available at `http://www.ics.forth.gr/cvrl/fire`.

Journal papers

- C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, A.A. Argyros, "FIRE: Fundus Image Registration Dataset", *Journal for Modeling in Ophthalmology*, vol. 1, no. 4, pp. 16-28, Jul. 2017.

- C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, A.A. Argyros, "Retinal Image Registration under the Assumption of a Spherical Eye", *Computerized Medical Imaging and Graphics*, Volume 55, January 2017, Pages 95-105. `DOI: 10.1016/j.compmedimag.2016.06.006`

[1] September 4, 2017
[2] `https://scholar.google.gr/citations?user=a_p0q64AAAAJ`

Conference papers

- C. Hernandez-Matas, X. Zabulis, A.A. Argyros, "An Experimental Evaluation of the Accuracy of Keypoints-based Retinal Image Registration", *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 377-381, Jeju Island, July 11-15, 2017

- C. Hernandez-Matas, X. Zabulis, A.A. Argyros, "Retinal Image Registration Through Simultaneous Camera Pose and Eye Shape Estimation", *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3247-3251, Orlando, August 16-20, 2016. `DOI: 10.1109/EMBC.2016.7591421`

- C. Hernandez-Matas, X. Zabulis, A.A. Argyros, "Retinal Image Registration Based on Keypoint Correspondences, Spherical Eye Modeling and Camera Pose Estimation", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5650-5654, Milan, August 25-29, 2015. `DOI: 10.1109/EMBC.2015.7319674`

- C. Hernandez-Matas, X. Zabulis, "Super Resolution for Fundoscopy Based on 3D Image Registration", *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6332-6338, Chicago, August 26-30, 2014. `DOI: 10.1109/EMBC.2014.6945077`

11.1.2 Non thesis-related publications

Additional work by the author in retinal imaging performed within the REVAMMAD project, but not reflected in this thesis, has been published as follows:

Journal papers

- S. Crespo-Garcia, N. Reichhart, C. Hernandez-Matas, X. Zabulis, N. Kociok, C. Brockmann, A.M. Joussen, O. Strauss, "In Vivo Analysis of the Time and Spatial Activation Pattern of Microglia in the Retina Following Laser-Induced Choroidal Neovascularization", *Experimental Eye Research*, Volume 139, October 2015, Pages 13-21. `DOI: 10.1016/j.exer.2015.07.012`

Extended abstracts

- S. Crespo-Garcia, N. Reichart, C. Hernandez-Matas, X. Zabulis, N. Kociok, A.M. Joussen, O. Strauss, "Image Processing Allows Reliable In Vivo Detection and Quantification of Retinal Microglia Following Laser-Induced Choroidal Neovascularization", *Investigative Ophthalmology & Visual Science*, vol. 56, issue 7, pp. 425, 2015

Part IV

Appendix and bibliography

Appendix A

Registration methods

This appendix aims to give an overview of Retinal Image Registration (RIR) methods in chronological order.

A.1 Peli et al. (1987) [43] presented a local registration method in the spatial domain that performs linear registration of intra-modal retinal images. This method utilizes a modified version of feature-based Sequential Similarity Detection in the vessel locations to find estimate the similarity of the reference and test image after transformation.

A.2 Cicediyan et al. (1992) [42] presented a global registration method in the frequency domain that performs linear transformations of intra-modal retinal images. They calculate the two dimensional FFT of each image and calculate its magnitude. They then sample a semi-circular ring of the Fourier magnitude limited by application-specific frequency cutoffs onto a log-polar grid in each image. Afterwards they zero-extend the log-polar Fourier magnitudes in the logarithm-of-radius coordinate direction only, before performing 2D cyclic cross-correlation using an FFT-based algorithm and determine the peak of the cross-correlation function and the corresponding rotation angle and scale factor. Then they appropriately rotate and scale one of the original images. This method is affected by textureless areas, has a high possibility of getting stuck in local minima. Only translation, rotation and scaling is assumed. It is a slow method.

A.3 Matsopoulos et al. (1999) [44] presented a local registration method in the spatial domain that performs affine and projective registration of intra- and cross-modal retinal images. It pre-processes the images by performing vessel enhancement and border suppression before proceeding to segment the vessels. Then they employ genetic algorithms to find the best matching transformation between the vessel trees.

A.4 Laliberté et al. (2003) [23] presented a local registration method in the spatial domain that performs linear and affine registration of intra- and cross-modal retinal images. Vessel bifurcations are extracted and matched and an iterative registration is performed, removing a pair of matched bifurcations after each iteration.

A.5 Stewart et al. (2003) [45] presented the Dual-Bootstrap Iterative Closest Point (DB-ICP) algorithm, a local registration method in the spatial domain that performs linear, affine or quadratic registration of intra- and cross-modal retinal images. The algorithm first extracts vessel centerlines, as well as bifurcations and crossovers via detection of elongated structures. Registration is initialized by matching bifurcations and estimating translation, rotation and scaling for the strongest bifurcation match. Afterwards, an iterative registration estimates the transformation parameters by calculating "point-to-line" distances from bifurcations in an image to vessel contours in the other as error metric. The area of points to match in the image grows after each iteration until convergence is obtained.

A.6 Ryan et al. (2004) [9] presented a local registration method in the spatial domain that performs affine or quadratic registration of intra- and cross-modal retinal images. They find vessel bifurcations on both images, and match them creating clusters of control points using a nearest neighbor algorithm with Euclidean distance. The clusters are matched utilizing expectation maximization. A suitable transformation is calculated using linear regression. Missing pixel values are interpolated using a spline technique.

A.7 Matsopoulos et al. (2004) [46] presented a local registration method in the spacial domain that performs affine registration of intra- and cross-modal retinal images. Vessel centerline detection and extraction of bifurcation points is performed. Automatic correspondence of bifurcations in the two images using self organizing maps are done. Then affine transform parameters are extracted.

A.8 Chanwimaluang et al. (2006) [47] presented a local method in the spatial domain that performs affine or quadratic registration of intra-modal retinal images. It extracts vessel bifurcations and match them combining mutual information and a feature method. Lens distortion is removed using the planar pattern calibration method proposed in [141]. They perform an initial retinal registration using affine factorization. Then affine bundle adjustment using a sparse Levenberg-Marquardt.

A.9 Choe et al. (2006) [12], [48] presented a local method in the spatial domain that performs affine registration of intra- and cross-modal images. They detect and match vessel bifurcations. Then estimate the fundamental matrix using plane+parallax algorithm from Kumar et al. [75] and use it to perform image rectification. Then they do stereo matching via mutual information. Finally, registration is performed.

A.10 Yang et al. (2007) [49] presented Generalized Dual Bootstrap-Iterative Closest Point (GDB-ICP), a local registration method in the spatial domain that performs linear, affine or quadratic registration of intra- and cross-modal retinal images and that improves on top of DB-ICP [45]. Scale-Invariant Feature Transform (SIFT) keypoint are matched and the strongest match is used to calculate an initial translation, rotation and scaling transformation. Afterwards, they use face and corner points to iteratively register the image pair utilizing quadratic transformation. This is a very widespread registration method, and is used as a standard for comparisons, as it offers good registration for the image pairs that is able to register. However, it struggles to successfully register challenging image pairs.

A.11 Lin et al. (2008) [11] presented a local registration method in the spatial domain that performs projective registration of intra- and cross-modal retinal images. They pre-process the images by enhancing their contrast. Then SIFT features are extracted and matched conventionally. Finally, chained registration is used to optimize the resulting registered image. After this 2D registration is performed, this algorithm adds as an advantage the estimation of the 3D structure of the eye, reconstructed from the two images registered, as seen in Section 4.4.1.

A.12 Chaudhry et al. (2008) [50] presented a local registration method in the spatial domain that performs affine registration of intra-modal retinal images. They enhance the contrast of the images using histogram equalization. Then the vascular tree is extracted by Gaussian filtering and black top-hat. Bifurcation points are extracted by counting the amount of distinct branches associated with a pixel. Then shape matching by calculating shape descriptors around each bifurcation point (as histograms) is performed. Finally, the affine transformation is calculated and applied.

A.13 Tsai et al. (2010) [36] presented Edge-Driven Dual-Bootstrap Iterative Closest Point (ED-DB-ICP), a local registration method in the spatial domain that performs linear, affine or quadratic registration of intra- and cross-modal retinal images and that improves on top of GDB-ICP. The main difference with its predecessor is utilizing corner and edge points instead of bifurcations, as well as extracting Difference of Gaussian keypoints from gradient-magnitude images.

A.14 Chen et al. (2010) [37] presented a local registration method in the spatial domain that performs linear, affine or quadratic registration of intra- and cross-modal retinal images. Corner points are selected and Partial Intensity Invariant Feature Descriptor (PIIFD) are extracted. The PIIFD of both images are matched bilaterally. Incorrect matches are removed and inaccurate matches are refined. An adaptive transformation is used to register the image pairs.

A.15 Deng et al. (2010) [51] presented a local registration method in the spatial domain that performs quadratic registration of intra- and cross-modal retinal images. Retinal vessels are automatically detected and represented as vascular structure graphs. A graph matching is then performed to find global correspondences between vascular bifurcations. Finally, a revised ICP algorithm with quadratic transformation model is used at fine level to register vessel shape models.

A.16 Perez-Rovira et al. (2011) [52] presented Robust Efficient Registration via Bifurcations and Elongated Elements (RERBEE), a local registration method in the spatial domain that performs quadratic registration of intra-modal retinal images. It performs vessel segmentation and Bifurcations and Elongated Elements Structure (BEES) extraction. BEES used to generate dense field of displacements. And finally, the transformation is applied.

A.17 Zheng et al. (2011) [53] presented a local registration method in the spatial domain that performs linear, affine, projective or quadratic registration of intra- and cross-modal retinal images. They extract salient feature regions using a local feature descriptor that combines both gradient field distribution and corresponding geometric distribution. Then perform coarse region pairs matching by calculating similarity measure between every pair of regions. Using a nearest neighbor, a cluster with the most pairs is calculated. Registration is performed using the cluster with the most pairs.

A.18 Troglio et al. (2011) [24], [54] presented a local registration in the spatial domains that performs affine registration of intra-modal retinal images. Firstly the vessel tree for the images is extracted. A genetic algorithm is then utilized to estimate the transformation between the vessel trees.

A.19 Gharabaghi et al. (2012) [25] presented a local registration method in the spatial domain that performs affine registration of intra-modal retinal images. Harris corners are used to detect blood vessels, and then close boundary regions are extracted by thresholding the area. The centers of gravity of the closed-boundary regions are used as keypoints. The features used are the first four affine moment invariants for each closed-boundary. The features extracted are used to find three pairs of most likely corresponding regions, forming a triangle. The angles of the triangles in both images are compared to validate the matching. Having the triangles, the parameters for the transformation are computed using the least square technique.

A.20 Ghassabi et al. (2013) [55] presented a local registration method in the spatial domain that performs quadratic registration of intra- and cross-modal retinal images. UR-SIFT is used to extract feature points, and their descriptors are calculated utilizing PIIFD. The points are matched conventionally, and the transformation parameters are calculated via a second-order polynomial transformation.

A.21 Reel et al. (2013) [26] presented a global registration method in the spatial domain that performs affine registration of intra- and cross-modal retinal images. The image information is rearranged into vector forms for a given neighborhood radius r, so that the spatial and intensity information is preserved. The first few principal components of the reference and sensed images are then iteratively computed by Expectation Maximization for Principal Component Analysis. The last step involves a mutual information computation from the data obtained in the previous step. It is computationally complex, not very fast and does not take into account 3D shape.

A.22 Legg et al. (2013) [56] presented a global registration method in the spatial domain that performs linear registration of intra- and cross-modal retinal images. The authors compare multiple variations of Mutual Information-based methods.

A.23 Bathina (2013) [27] presented a local registration method in the spatial domain that performs affine, projective or quadratic registration of intra- and cross-modal retinal images. It performs vessel enhancement based on a vesselness measure and landmark detection using Curvature Dispersion Measure. Each landmark is computer via radon based descriptor and matched using bilateral matching. False matches are rejected. The transformation is estimated using M-estimators Sample and Consensus (MSAC) and refined using normalized cross-correlation. The final transformation is then computed using M-estimators.

A.24 Hernandez-Matas et al. (2014) [6] presented a local registration method in the spatial domain that performs projective registration of intra-modal retinal images. Speeded Up Robust Features (SURF) features in the green channel of images are extracted and matched. Homography is estimated and the test image warped.

A.25 Chen et al. (2014) [57] presented a local registration method in the spatial domain that performs linear or affine registration of intra- and cross-modal retinal images. Vessels and bifurcations are extracted. Bifurcations in both images are matched using structure matching and taking into account each bifurcation's branches and angles. Then the affine transformation is calculated.

A.26 Adal et al. (2014) [28] presented a global registration method in the spatial domain that performs linear, affine or quadratic registration of intra-modal retinal images. First, image normalization is performed. The normalized images are then aligned based on a vasculature-weighted mean squared difference (MSD) similarity metric.

A.27 Lee et al. (2015) [58] presented a local registration method in the spatial domain that performs affine registration of intra- and cross-modal retinal images. Geometric corners are extracted, and a new descriptor denominated low-dimensional step pattern analysis (LoSPA) is utilized to calculate descriptors. Keypoints are matched using Euclidean distances and the affine transformation is calculated.

A.28 Wang et al. (2015) [29] presented a local registration method in the spatial domain that performs linear, affine or quadratic registration of intra- and cross-modal retinal images. Local feature points are located by a SURF detector. Descriptors are extracted using PIIFD. Feature matching and mismatch removal is performed using Robust Point Matching (RPM). Finally, the transformation is estimated using weighted least-squares.

A.29 Ghassabi et al. (2015) [59] presented a local registration method in the spatial domain that performs linear, affine or quadratic registration of intra-modal retinal images. A region detector based on a robust watershed segmentation obtains closed-boundary regions within a clean vascular structure map. The regions are approximated by convex polygons, and boundary descriptors are achieved to match them. Correspondences determine the parameters of geometric transformation between input images.

A.30 Liu et al. (2016) [60] presented a local registration method in the spatial domain that performs affine registration of intra- and cross-modal retinal images. A Gaussian mixture model is fitted to one point set, such that both the centers and local features of the Gaussian densities are constrained to coincide with the other point set. The problem is solved under a unified maximum-likelihood framework together with an iterative Expectation Maximization algorithm initialized by the confident feature correspondences.

A.31 Saha et al. (2016) [61] presented a local registration method in the spatial domain that performs quadratic registration of intra-modal retinal images. A preliminary registration is performed relying on SURF keypoint matching. Then Binary Robust Independent Elementary Features (BRIEF) are computed around bifurcation points which are then matched based on Hamming distance.

Bibliography

[1] A Grosso, F Veglio, M Porta, F. M. Grignolo, and T. Y. Wong, "Hypertensive retinopathy revisited: some answers, more questions", *British Journal of Ophthalmology*, vol. 89, no. 12, pp. 1646–1654, 2005. DOI: 10.1136/bjo.2005.072546.

[2] R. P. Danis and M. D. Davis, "Proliferative diabetic retinopathy", in *Diabetic Retinopathy*. Totowa, NJ: Humana Press, 2008, pp. 29–65, ISBN: 978-1-59745-563-3. DOI: 10.1007/978-1-59745-563-3_2.

[3] M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal Imaging and Image Analysis", *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010, ISSN: 1937-3333. DOI: 10.1109/RBME.2010.2084567.

[4] N. Meitav and E. N. Ribak, "Improving retinal image resolution with iterative weighted shift-and-add", *Journal of the Optical Society of America A*, vol. 28, no. 7, pp. 1395–1402, Jul. 2011. DOI: 10.1364/JOSAA.28.001395.

[5] G. Molodij, E. Ribak, M. Glanc, and G. Chenegros, "Enhancing retinal images by extracting structural information", *Optics Communications*, vol. 313, pp. 321 –328, 2014, ISSN: 0030-4018. DOI: 10.1016/j.optcom.2013.10.011.

[6] C. Hernandez-Matas and X. Zabulis, "Super resolution for fundoscopy based on 3D image registration", in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2014, pp. 6332–6338. DOI: 10.1109/EMBC.2014.6945077.

[7] R Hubbard L Brothers, W King, L Clegg, R Klein, L Cooper, A Sharrett, M Davis, and J Cai, "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.

[8] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 412–419, Mar. 2002, ISSN: 0162-8828. DOI: 10.1109/34.990145.

[9] N. Ryan, C. Heneghan, and P. de Chazal, "Registration of digital retinal images using landmark correspondence by expectation maximization", *Image and Vision Computing*, vol. 22, no. 11, pp. 883 –898, 2004, ISSN: 0262-8856. DOI: 10.1016/j.imavis.2004.04.004.

[10] P. C. Cattin, H. Bay, L. Van Gool, and G. Székely, "Retina mosaicing using local features", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part II*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 185–192, ISBN: 978-3-540-44728-3. DOI: 10.1007/11866763_23.

[11] Y. Lin and G. Medioni, "Retinal image registration from 2D to 3D", in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587705.

[12] T. E. Choe, I. Cohen, G. Medioni, A. C. Walsh, and S. R. Sadda, "Evaluation of 3-D shape reconstruction of retinal fundus", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 134–141, ISBN: 978-3-540-44708-5. DOI: 10.1007/11866565_17.

[13] T. Chanwimaluang, G. Fan, G. G. Yen, and S. R. Fransen, "3-D retinal curvature estimation", *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 997–1005, Nov. 2009, ISSN: 1089-7771. DOI: 10.1109/TITB.2009.2027014.

[14] Li Tang, M. K. Garvin, Kyungmoo Lee, W. L. W. Alward, Y. H. Kwon, and M. D. Abramoff, "Robust Multiscale Stereo Matching from Fundus Images with Radiometric Differences", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2245–2258, Nov. 2011, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2011.69.

[15] 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 transactions on bio-medical engineering*, vol. 54, no. 8, pp. 1436–45, Aug. 2007, ISSN: 0018-9294. DOI: 10.1109/TBME.2007.900807.

[16] G. Troglio, M. Alberti, J. A. Benediksson, G. Moser, S. B. Serpico, and E. Stefánsson, "Unsupervised change-detection in retinal images by a multiple-classifier approach", in *Multiple Classifier Systems: 9th International Workshop, MCS 2010, Cairo, Egypt, April 7-9, 2010. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 94–103, ISBN: 978-3-642-12127-2. DOI: 10.1007/978-3-642-12127-2_10.

[17] World Health Organization, *A global brief on Hypertension - World Health Day 2013*. World Health Organization, 2013, [Accessed May 25, 2017]. [Online]. Available: `http://apps.who.int/iris/bitstream/10665/79059/1/WHO_DCO_WHD_2013.2_eng.pdf`.

[18] ——, *Global Report on Diabetes*. World Health Organization, 2016, [Accessed May 25, 2017]. [Online]. Available: `http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf`.

[19] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010", *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012, ISSN: 0007-1161. DOI: `10.1136/bjophthalmol-2011-300539`.

[20] H. Gross, F. Blechinger, W. Singer, and B. Achtner, *Handbook of Optical Systems, Vol. 4: Survey of Optical Instruments*. Wiley-VCH Verlag GmbH & Co. KGaA, 2008, p. 1092, ISBN: 978-3-527-40380-5. DOI: `10.1002/3527606688.fmatter`.

[21] W Ogilvy, "Ophthalmoscopy", *The Australasian Journal of Optometry*, vol. 5, no. 7, pp. 143–148, 1923. DOI: `10.1111/j.1444-0938.1923.tb01720.x`.

[22] H. R. Hildreth, "Red-free opthalmoscopy", *Clinical and Experimental Optometry*, vol. 18, no. 4, pp. 165–166, 1935. DOI: `10.1111/j.1444-0938.1935.tb05701.x`.

[23] F. Laliberte, L. Gagnon, and Y. Sheng, "Registration and fusion of retinal images-an evaluation study", *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 661–673, May 2003, ISSN: 0278-0062. DOI: `10.1109/TMI.2003.812263`.

[24] G. Troglio, J. A. Benediktsson, G. Moser, S. B. Serpico, and E. Stefansson, "Unsupervised change detection in multitemporal images of the human retina", in *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Volume 1*. Boston, MA: Springer US, 2011, pp. 309–337, ISBN: 978-1-4419-8195-0. DOI: `10.1007/978-1-4419-8195-0_11`.

[25] S. Gharabaghi, S. Daneshvar, and M. H. Sedaaghi, "Retinal Image Registration Using Geometrical Features", *Journal of Digital Imaging*, vol. 26, no. 2, pp. 248–258, 2013, ISSN: 1618-727X. DOI: `10.1007/s10278-012-9501-7`.

[26] P. S. Reel, L. S. Dooley, K. C. P. Wong, and A. Börner, "Robust retinal image registration using expectation maximisation with mutual information", in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1118–1122. DOI: `10.1109/ICASSP.2013.6637824`.

[27] Y. B. Bathina, "Robust Registration of Retinal Images", PhD thesis, International Institute of Information Technology Hyderabad, 2013.

[28] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. J. van Vliet, "A Hierarchical Coarse-to-Fine Approach for Fundus Image Registration", *Biomedical Image Registration. WBIR 2014. Lecture Notes in Computer Science*, vol. 8545, pp. 93–102, 2014. DOI: 10.1007/978-3-319-08554-8_10.

[29] G. Wang, Z. Wang, Y. Chen, and W. Zhao, "Robust point matching method for multimodal retinal image registration", *Biomedical Signal Processing and Control*, vol. 19, pp. 68 –76, 2015, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2015.03.004.

[30] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration based on keypoint correspondences, spherical eye modeling and camera pose estimation", in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015, pp. 5650–5654. DOI: 10.1109/EMBC.2015.7319674.

[31] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. A. Argyros, "Retinal image registration under the assumption of a spherical eye", *Computerized Medical Imaging and Graphics*, vol. 55, pp. 95 –105, 2017, Special Issue on Ophthalmic Medical Image Analysis, ISSN: 0895-6111. DOI: 10.1016/j.compmedimag.2016.06.006.

[32] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "Retinal image registration through simultaneous camera pose and eye shape estimation", in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, pp. 3247–3251. DOI: 10.1109/EMBC.2016.7591421.

[33] G. Lu and B. Fei, "Medical hyperspectral imaging: A review", *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010 901, 2014. DOI: 10.1117/1.JBO.19.1.010901.

[34] B. F. Hochheimer, "Angiography of the retina with indocyanine green", *Archives of Ophthalmology*, vol. 86, no. 5, pp. 564–565, 1971. DOI: 10.1001/archopht.1971.01000010566014.

[35] N. S. Claxton, T. J. Fellers, and M. W. Davidson, "Microscopy, confocal", in *Encyclopedia of Medical Devices and Instrumentation*. John Wiley & Sons, Inc., 2006, ISBN: 9780471732877. DOI: 10.1002/0471732877.emd291.

[36] C. L. Tsai, C. Y. Li, G. Yang, and K. S. Lin, "The Edge-Driven Dual-Bootstrap Iterative Closest Point Algorithm for Registration of Multimodal Fluorescein Angiogram Sequence", *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 636–649, Mar. 2010, ISSN: 0278-0062. DOI: 10.1109/TMI.2009.2030324.

[37] 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 Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1707–1718, Jul. 2010, ISSN: 0018-9294. DOI: 10.1109/TBME.2010.2042169.

[38] A. Sakamoto, M. Hangai, and N. Yoshimura, "Spectral-Domain Optical Coherence Tomography with multiple B-scan averaging for enhanced imaging of retinal diseases", *Ophthalmology*, vol. 115, no. 6, 1071 –1078.e7, 2008, ISSN: 0161-6420. DOI: 10.1016/j.ophtha.2007.09.001.

[39] B. Zitová and J. Flusser, "Image registration methods: A survey", *Image and Vision Computing*, vol. 21, no. 11, pp. 977 –1000, 2003, ISSN: 0262-8856. DOI: 10.1016/S0262-8856(03)00137-9.

[40] D. Rueckert and J. A. Schnabel, "Medical image registration", in *Biomedical Image Processing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 131–154, ISBN: 978-3-642-15816-2. DOI: 10.1007/978-3-642-15816-2_5.

[41] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey", in *Image and Signal Processing: 4th International Conference, ICISP 2010, Trois-Rivières, QC, Canada, June 30-July 2, 2010. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 103–112, ISBN: 978-3-642-13681-8. DOI: 10.1007/978-3-642-13681-8_13.

[42] A. V. Cideciyan, S. G. Jacobson, C. M. Kemp, R. W. Knighton, and J. H. Nagel, "Registration of High Resolution Images of the Retina", *SPIE Medical Imaging*, vol. 1652, pp. 310–322, Jun. 1992. DOI: 10.1117/12.59439.

[43] E. Peli, R. A. Augliere, and G. T. Timberlake, "Feature-based registration of retinal images", *IEEE Transactions on Medical Imaging*, vol. 6, no. 3, pp. 272–278, Sep. 1987, ISSN: 0278-0062. DOI: 10.1109/TMI.1987.4307837.

[44] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques", *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, no. 1, pp. 47–60, Mar. 1999, ISSN: 1089-7771. DOI: 10.1109/4233.748975.

[45] C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration", *IEEE Transactions on Medical Imaging*, vol. 22, no. 11, pp. 1379–1394, Nov. 2003, ISSN: 0278-0062. DOI: 10.1109/TMI.2003.819276.

[46] G. K. Matsopoulos, P. A. Asvestas, N. A. Mouravliansky, and K. 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, ISSN: 0278-0062. DOI: 10.1109/TMI.2004.836547.

[47] T. Chanwimaluang, "Advanced retinal imaging: feature extraction, 2-D registration, and 3-D reconstruction", PhD thesis, Oklahoma State University, 2006.

[48] T. E. Choe, G. Medioni, I. Cohen, A. C. Walsh, and S. R. Sadda, "2-D registration and 3-D shape inference of the retinal fundus from fluorescein images", *Medical Image Analysis*, vol. 12, no. 2, pp. 174 –190, 2008, ISSN: 1361-8415. DOI: `http://dx.doi.org/10.1016/j.media.2007.10.002`.

[49] G. Yang, C. V. Stewart, M. Sofka, and C. L. Tsai, "Registration of Challenging Image Pairs: Initialization, Estimation, and Decision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1973–1989, Nov. 2007, ISSN: 0162-8828. DOI: `10.1109/TPAMI.2007.1116`.

[50] A. R. Chaudhry and J. C. Klein, "Ophthalmologic Image Registration based on shape-context: Application to Fundus Autofluorescence (FAF) images", *Visualization, Imaging, and Image Processing*, pp. 1–7, 2008.

[51] K. Deng, J. Tian, J. Zheng, X. Zhang, X. Dai, and M. Xu, "Retinal fundus image registration via vascular structure graph matching", *Journal of Biomedical Imaging*, vol. 2010, 14:1–14:13, Jan. 2010, ISSN: 1687-4188. DOI: `10.1155/2010/906067`.

[52] 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 Transactions on Medical Imaging*, vol. 31, no. 1, pp. 140–150, Jan. 2012, ISSN: 0278-0062. DOI: `10.1109/TMI.2011.2167517`.

[53] 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. 221–232, Mar. 2011, ISSN: 1089-7771. DOI: `10.1109/TITB.2010.2091145`.

[54] G. Troglio, A. Nappo, J. A. Benediktsson, G. Moser, S. B. Serpico, and E. Stefansson, "Automatic change detection of retinal images", in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany: Vol. 25/11 Biomedical Engineering for Audiology, Ophthalmology, Emergency & Dental Medicine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 281–284, ISBN: 978-3-642-03891-4. DOI: `10.1007/978-3-642-03891-4_75`.

[55] Z. Ghassabi, J. Shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, "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, p. 25, 2013, ISSN: 1687-5281. DOI: 10.1186/1687-5281-2013-25.

[56] P. Legg, P. Rosin, D. Marshall, and J. Morgan, "Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation", *Computerized Medical Imaging and Graphics*, vol. 37, no. 7–8, pp. 597 –606, 2013, ISSN: 0895-6111. DOI: 10.1016/j.compmedimag.2013.08.004.

[57] L. Chen, X. Huang, and J. Tian, "Retinal image registration using topological vascular tree segmentation and bifurcation structures", *Biomedical Signal Processing and Control*, vol. 16, pp. 22 –31, 2015, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2014.10.009.

[58] J. A. Lee, J. Cheng, B. H. Lee, E. P. Ong, G. Xu, D. W. K. Wong, J. Liu, A. Laude, and T. H. Lim, "A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration", in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1046–1053. DOI: 10.1109/CVPR.2015.7298707.

[59] Z. Ghassabi, J. Shanbehzadeh, and A. Mohammadzadeh, "A structure-based region detector for high-resolution retinal fundus image registration", *Biomedical Signal Processing and Control*, vol. 23, pp. 52 –61, 2016, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2015.08.005.

[60] C. Liu, J. Ma, Y. Ma, and J. Huang, "Retinal image registration via feature-guided gaussian mixture model", *Journal of the Optical Society of America A*, vol. 33, no. 7, pp. 1267–1276, Jul. 2016. DOI: 10.1364/JOSAA.33.001267.

[61] S. K. Saha, D. Xiao, S. Frost, and Y. Kanagasingam, "A two-step approach for longitudinal registration of retinal images", *Journal of Medical Systems*, vol. 40, no. 12, p. 277, 2016, ISSN: 1573-689X. DOI: 10.1007/s10916-016-0640-0.

[62] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, "An Experimental Evaluation of the Accuracy of Keypoints-based Retinal Image Registration", in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 377–381.

[63] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey", *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, Mar. 2005, ISSN: 1057-7149. DOI: 10.1109/TIP.2004.838698.

[64] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution", *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004, ISSN: 1057-7149. DOI: 10.1109/TIP.2004.834669.

[65] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using subpixel displacements", in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*, Jun. 1988, pp. 742–746. DOI: 10.1109/CVPR.1988.196317.

[66] M. Irani and S. Peleg, "Improving resolution by image registration", *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231 –239, 1991, ISSN: 1049-9652. DOI: 10.1016/1049-9652(91)90045-L.

[67] L. Lucchese and G. M. Cortelazzo, "A noise-robust frequency domain technique for estimating planar roto-translations", *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1769–1786, Jun. 2000, ISSN: 1053-587X. DOI: 10.1109/78.845934.

[68] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution", *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 071 459, 2006, ISSN: 1687-6180. DOI: 10.1155/ASP/2006/71459.

[69] T. Köhler, A. Brost, K. Mogalle, Q. Zhang, C. Köhler, G. Michelson, J. Hornegger, and R. P. Tornow, "Multi-frame super-resolution with quality self-assessment for retinal fundus videos", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part I*. Cham: Springer International Publishing, 2014, pp. 650–657, ISBN: 978-3-319-10404-1. DOI: 10.1007/978-3-319-10404-1_81.

[70] C. C. Hsu and C. W. Lin, "Image super-resolution via feature-based affine transform", in *2011 IEEE 13th International Workshop on Multimedia Signal Processing*, Oct. 2011, pp. 1–5. DOI: 10.1109/MMSP.2011.6093845.

[71] F. Laliberte, L. Gagnon, and Y. Sheng, "Three-dimensional visualization of human fundus from a sequence of angiograms", *Proceedings SPIE*, vol. 5664, pp. 412–420, 2005. DOI: 10.1117/12.583379.

[72] A. Cheriyan, B. Menon, and K. Narayanankutty, "3D reconstruction of human retina from a single fundus image", *Proceedings of the 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2011*, vol. 1, pp. 252–257, 2011.

[73] M. Martinez-Perez and A Espinosa-Romero, “3D Reconstruction of Retinal Blood Vessels from Two Views.”, *The Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2004.

[74] D. Liu, N. B. Wood, X. Y. Xu, N. Witt, A. D. Hughes, and T. SAMcG, “3D reconstruction of the retinal arterial tree using subject-specific fundus images”, in *Advances in Computational Vision and Medical Image Processing: Methods and Applications*. Dordrecht: Springer Netherlands, 2009, pp. 187–201, ISBN: 978-1-4020-9086-8. DOI: 10.1007/978-1-4020-9086-8_11.

[75] R. Kumar, P. Anandan, and K. Hanna, “Direct recovery of shape from multiple views: A parallax based approach”, in *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, Oct. 1994, 685–688 vol.1. DOI: 10.1109/ICPR.1994.576402.

[76] E Kruppa, “Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung”, *Sitzungsberichte der Mathematisch Naturwissenschaftlichen Kaiserlichen Akademie der Wissenschaften*, vol. 122, pp. 1939–1948, 1913.

[77] A. M. Broehan, T. Rudolph, C. A. Amstutz, and J. H. Kowal, “Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system”, *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2816–2824, Oct. 2011, ISSN: 0018-9294. DOI: 10.1109/TBME.2011.2159860.

[78] S. Crespo-Garcia, N. Reichhart, C. Hernandez-Matas, X. Zabulis, N. Kociok, C. Brockmann, A. M. Joussen, and O. Strauß, “Invivo analysis of the time and spatial activation pattern of microglia in the retina following laser-induced choroidal neovascularization”, *Experimental Eye Research*, vol. 139, pp. 13 –21, 2015, ISSN: 0014-4835. DOI: 10.1016/j.exer.2015.07.012.

[79] T. S. Huang and Y. P. Hsu, “Image sequence enhancement”, in *Image Sequence Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 289–309, ISBN: 978-3-642-87037-8. DOI: 10.1007/978-3-642-87037-8_4.

[80] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. New York, NY, USA: John Wiley & Sons, Inc., 1987, ISBN: 0-471-85233-3.

[81] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004, ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000029664.99615.94.

[82] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Speeded-Up Robust Features (SURF)”, *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 –359, 2008, Similarity Matching in Computer Vision and Multimedia, ISSN: 1077-3142. DOI: 10.1016/j.cviu.2007.09.014.

[83] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering", in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 130–137, ISBN: 978-3-540-49563-5. DOI: 10.1007/BFb0056195.

[84] 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, ISSN: 0001-0782. DOI: 10.1145/357994.358023.

[85] M. A. Fischler and R. C. 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. 381–395, Jun. 1981, ISSN: 0001-0782. DOI: 10.1145/358669.358692.

[86] M. Lourakis and X. Zabulis, "Model-based pose estimation for rigid objects", in *Computer Vision Systems: 9th International Conference, ICVS 2013, St. Petersburg, Russia, July 16-18, 2013. Proceedings*, M. Chen, B. Leibe, and B. Neumann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 83–92, ISBN: 978-3-642-39402-7. DOI: 10.1007/978-3-642-39402-7_9.

[87] J. Kennedy and R. Eberhart, "Particle swarm optimization", in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, 1995, 1942–1948 vol.4. DOI: 10.1109/ICNN.1995.488968.

[88] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization", *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007. DOI: 10.1007/s11721-007-0002-0.

[89] X. Zabulis, M. Lourakis, and P. Koutlemanis, "3d object pose refinement in range images", in *Proceedings of the 10th International Conference on Computer Vision Systems - Volume 9163*, ser. ICVS 2015, Copenhagen, Denmark: Springer-Verlag New York, Inc., 2015, pp. 263–274, ISBN: 978-3-319-20903-6. DOI: 10.1007/978-3-319-20904-3_25.

[90] R. Navarro, J. Santamaría, and J. Bescós, "Accommodation-dependent model of the human eye with aspherics", *Journal of the Optical Society of America A*, vol. 2, no. 8, pp. 1273–1280, Aug. 1985. DOI: 10.1364/JOSAA.2.001273.

[91] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Second edition. Cambridge University Press, 2004.

[92] 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 *2011 International Conference on Computer Vision*, Nov. 2011, pp. 2088–2095. DOI: 10.1109/ICCV.2011.6126483.

[93] N. Kyriazis and A. Argyros, "Scalable 3D Tracking of Multiple Interacting Objects", in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 3430–3437. DOI: 10.1109/CVPR.2014.438.

[94] I. Escudero-Sanz and R. Navarro, "Off-axis aberrations of a wide-angle schematic eye model", *Journal of the Optical Society of America A*, vol. 16, no. 8, pp. 1881–1891, Aug. 1999. DOI: 10.1364/JOSAA.16.001881.

[95] M. A. Lourakis and A. Argyros, "SBA: A Software Package for Generic Sparse Bundle Adjustment", *ACM Trans. Math. Software*, vol. 36, no. 1, pp. 1–30, 2009. DOI: http://doi.acm.org/10.1145/1486525.1486527.

[96] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012, ISSN: 0018-9294. DOI: 10.1109/TBME.2012.2205687.

[97] *CHASEDB1 Retinal Image Database*, [Accessed May 25, 2017]. [Online]. Available: https://blogs.kingston.ac.uk/retinal/chasedb1/.

[98] E. J. Carmona, M. Rincón, J. García-Feijoó, and J. M. Martínez-de-la Casa, "Identification of the Optic Nerve Head with Genetic Algorithms", *Artificial Intelligence in Medicine*, vol. 43, no. 3, pp. 243–259, Jul. 2008, ISSN: 0933-3657. DOI: 10.1016/j.artmed.2008.04.005.

[99] *DRIONS-DB: Digital Retinal Images for Optic Nerve Segmentation Database*, [Accessed May 25, 2017]. [Online]. Available: http://www.ia.uned.es/~ejcarmona/DRIONS-DB.html.

[100] J. Sivaswamy, S. R. Krishnadas, G. D. Joshi, M. Jain, and A. U. S. Tabish, "Drishti-GS: Retinal image dataset for optic nerve head(ONH) segmentation", in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2014, pp. 53–56. DOI: 10.1109/ISBI.2014.6867807.

[101] *Drishti-GS Dataset*, [Accessed May 25, 2017]. [Online]. Available: http://cvit.iiit.ac.in/projects/mip/drishti-gs.

[102] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina", *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004, ISSN: 0278-0062. DOI: 10.1109/TMI.2004.825627.

[103] *DRIVE: Digital Retinal Images for Vessel Extraction*, [Accessed May 25, 2017]. [Online]. Available: http://www.isi.uu.nl/Research/Databases/DRIVE/.

[104] J. Odstrčilík, J. Jan, J. Gazárek, and R. Kolář, "Improvement of Vessel Segmentation by Matched Filtering in Colour Retinal Images", in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany: Vol. 25/11 Biomedical Engineering for Audiology, Ophthalmology, Emergency & Dental Medicine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 327–330, ISBN: 978-3-642-03891-4. DOI: `10.1007/978-3-642-03891-4_87`.

[105] *HRF: High-Resolution Fundus Image Database*, [Accessed May 25, 2017]. [Online]. Available: `https://www5.cs.fau.de/research/data/fundus-images/`.

[106] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: the MESSIDOR database", *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014, ISSN: 1854-5165. DOI: `10.5566/ias.1155`.

[107] *MESSIDOR: Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology*, [Accessed May 25, 2017]. [Online]. Available: `http://www.adcis.net/en/Download-Third-Party/Messidor.html`.

[108] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation", *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 256–264, Feb. 2004, ISSN: 0278-0062. DOI: `10.1109/TMI.2003.823261`.

[109] *ONHSD: Optic Nerve Head Segmentation Dataset*, [Accessed May 25, 2017]. [Online]. Available: `http://reviewdb.lincoln.ac.uk/Image%20Datasets/ONHSD.aspx`.

[110] B. Al-Diri, A. Hunter, D. Steel, M. Habib, T. Hudaib, and S. Berry, "REVIEW - A reference data set for retinal vessel profiles", in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008, pp. 2262–2265. DOI: `10.1109/IEMBS.2008.4649647`.

[111] *REVIEW: Retinal Vessel Image set for Estimation of Widths*, [Accessed May 25, 2017]. [Online]. Available: `http://reviewdb.lincoln.ac.uk/REVIEWDB/REVIEWDB.aspx`.

[112] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, and J. Pietilä, *DIARETDB0: Evaluation Database and Methodology for Diabetic Retinopathy Algorithms.*

[113] *DIARETDB0: Standard Diabetic Retinopathy Database Calibration level 0*, [Accessed May 25, 2017]. [Online]. Available: `http://www2.it.lut.fi/project/imageret/diaretdb0/`.

[114] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A Raninen, R Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, *DIARETDB1 diabetic retinopathy database and evaluation protocol.*

[115] *DIARETDB1: Standard Diabetic Retinopathy Database Calibration level 1*, [Accessed May 25, 2017]. [Online]. Available: `http://www2.it.lut.fi/project/imageret/diaretdb1/`.

[116] E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Laÿ, and A. Chabouis, “TeleOphta: Machine learning and image processing methods for teleophthalmology”, *IRBM*, vol. 34, no. 2, pp. 196 –203, 2013, Special issue: ANR TECSAN: Technologies for Health and Autonomy, ISSN: 1959-0318. DOI: `10.1016/j.irbm.2013.01.010`.

[117] *e-ophtha: A Color Fundus Image Database*, [Accessed May 25, 2017]. [Online]. Available: `http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html`.

[118] M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. van Ginneken, J. C. Folk, and M. D. Abramoff, “Automated Measurement of the Arteriolar-to-Venular Width Ratio in Digital Color Fundus Photographs”, *IEEE Transactions on Medical Imaging*, vol. 30, no. 11, pp. 1941–1950, Nov. 2011, ISSN: 0278-0062. DOI: `10.1109/TMI.2011.2159619`.

[119] *INSPIRE: Iowa Normative Set for Processing Images of the Retina*, [Accessed May 25, 2017]. [Online]. Available: `http://medicine.uiowa.edu/eye/inspire-datasets`.

[120] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. D. Abramoff, “Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs”, *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 185–195, Jan. 2010, ISSN: 0278-0062. DOI: `10.1109/TMI.2009.2033909`.

[121] *ROC: Retinopathy Online Challenge Database*, [Accessed May 25, 2017]. [Online]. Available: `http://webeye.ophth.uiowa.edu/ROC/`.

[122] A. Hoover and M. Goldbaum, “Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels”, *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 951–958, Aug. 2003, ISSN: 0278-0062. DOI: `10.1109/TMI.2003.815900`.

[123] *STARE: Structured Analysis of the Retina*, [Accessed May 25, 2017]. [Online]. Available: `http://cecas.clemson.edu/~ahoover/stare/index.html`.

[124] S. G. Vázquez, B. Cancela, N. Barreira, M. G. Penedo, M. Rodríguez-Blanco, M. Pena Seijo, G. C. de Tuero, M. A. Barceló, and M. Saez, "Improving retinal artery and vein classification by means of a minimal path approach", *Machine Vision and Applications*, vol. 24, no. 5, pp. 919–930, 2013, ISSN: 1432-1769. DOI: 10.1007/s00138-012-0442-4.

[125] *VICAVR Database*, [Accessed May 25, 2017]. [Online]. Available: http://www.varpa.es/research/ophtalmology.html#databases.

[126] M. Ortega, M. G. Penedo, J. Rouco, N. Barreira, and M. J. Carreira, "Retinal Verification Using a Feature Points-based Biometric Pattern", *EURASIP Journal on Advances in Signal Processing - Special issue on recent advances in biometric systems: a signal processing perspective*, vol. 2009, 2:1–2:13, Jan. 2009, ISSN: 1110-8657. DOI: 10.1155/2009/235746.

[127] *VARIA*, [Accessed May 25, 2017]. [Online]. Available: http://www.varpa.es/research/biometrics.html#databases.

[128] 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*, vol. 1, no. 4, pp. 16–28, 2017.

[129] *FIRE: Fundus Image Registration Dataset.* [Online]. Available: http://www.ics.forth.gr/cvrl/fire.

[130] K. M. Adal, P. G. van Etten, J. P. Martinez, L. J. van Vliet, and K. A. Vermeer, "Accuracy asessment of intra- and intervisit fundus image registration for diabetic retinopathy screening", *Investigative Ophthalmology & Visual Science*, vol. 56, no. 3, pp. 1805–1812, 2015. DOI: 10.1167/iovs.14-15949.

[131] *RODREP: Rotterdam Ophthalmic Data Repository Longitudinal diabetic retinopathy screening data*, [Accessed May 25, 2017]. [Online]. Available: http://www.rodrep.com/longitudinal-diabetic-retinopathy-screening---description.html.

[132] M. Hernandez, G. Medioni, Z. Hu, and S. Sadda, "Multimodal Registration of Multiple Retinal Images Based on Line Structures", in *2015 IEEE Winter Conference on Applications of Computer Vision*, Jan. 2015, pp. 907–914. DOI: 10.1109/WACV.2015.125.

[133] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, "Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes", in *Computer Vision – ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I.* Berlin, Heidelberg:

Springer Berlin Heidelberg, 2013, pp. 548–562, ISBN: 978-3-642-37331-2. DOI: 10.1007/978-3-642-37331-2_42.

[134] B. Babenko, M. H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.226.

[135] T. Q. Pham, L. J. van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution", *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 083 268, 2006, ISSN: 1687-6180. DOI: 10.1155/ASP/2006/83268.

[136] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution", in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. 645–650. DOI: 10.1109/CVPR.2001.990535.

[137] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images", *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, Jan. 2006, ISSN: 1057-7149. DOI: 10.1109/TIP.2005.860336.

[138] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches", in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, Jun. 2005, pp. 510–517. DOI: 10.1109/CVPR.2005.235.

[139] S. Gauglitz, L. Foschini, M. Turk, and T. Höllerer, "Efficiently selecting spatially distributed keypoints for visual tracking", in *2011 18th IEEE International Conference on Image Processing*, Sep. 2011, pp. 1869–1872. DOI: 10.1109/ICIP.2011.6115832.

[140] M. Lourakis and R. Deriche, "Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3D measurements", *RR-3748, INRIA*, 1999.

[141] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, 1999, 666–673 vol.1. DOI: 10.1109/ICCV.1999.791289.