UFR DE MÉDECINE
ET DES PROFESSIONS PARAMÉDICALES
Université Clermont Auvergne

Année N° d'ordre

ECOLE DOCTORALE
DES SCIENCES POUR L'INGENIEUR

THÈSE

Présentée à l'Université Clermont Auvergne
pour l'obtention du grade de **DOCTEUR**
(Décret du 5 juillet 1984)

Spécialité
COMPUTER VISION

Soutenue le
10 Novembre 2017

Kristina PROKOPETC

Precise Mapping for Retinal Photocoagulation in SLIM (Slit-Lamp Image Mosaicing)

Examinatrice	Isabelle BARTHELEMY	PU-PH, Université Clermont Auvergne, France
Rapporteurs	Raphaël SZNITMAN	MCU, Université de Berne, Suisse
	Fabrice MERIAUDEAU	PU, Universiti Teknologi PETRONAS, Malaisie
Directeur de thèse	Adrien BARTOLI	PU, Université Clermont Auvergne, France
Invité	Benjamin WASSMER	Ingénieur R&D, Quantel Medical, France

Quantel Medical
11 Rue du Bois Joli
Cournon-d'Auvergne
Tel: +33 4 73 74 57 45

EnCoV, IP, UMR 6602 CNRS
Université Clermont Auvergne
28 place Henri Dunant, Clermont-Ferrand
Tel: +33 4 73 17 81 23

Abstract

This thesis arises from an agreement Convention Industrielle de Formation par la REcherche (CIFRE) between the Endoscopy and Computer Vision (EnCoV) research group at Université Clermont Auvergne and the company Quantel Medical (www.quantel-medical.fr), which specializes in the development of innovative ultrasound and laser products in ophthalmology. It presents a research work directed at the application of computer-aided diagnosis and treatment of retinal diseases with a use of the TrackScan industrial prototype developed at Quantel Medical. More specifically, it contributes to the problem of precise Slit-Lamp Image Mosaicing (SLIM) and automatic multi-modal registration of SLIM with Fluorescein Angiography (FA) to assist navigated pan-retinal photocoagulation. We address three different problems.

The first is a problem of accumulated registration errors in SLIM, namely the mosaicing drift. A common approach to image mosaicing is to compute transformations only between temporally consecutive images in a sequence and then to combine them to obtain the transformation between non-temporally consecutive views. Many existing algorithms follow this approach. Despite the low computational cost and the simplicity of such methods, due to its 'chaining' nature, alignment errors tend to accumulate, causing images to drift in the mosaic. We propose to use recent advances in key-frame Bundle Adjustment methods and present a drift reduction framework that is specifically designed for SLIM. We also introduce a new local refinement procedure.

Secondly, we tackle the problem of various types of light-related imaging artifacts common in SLIM, which significantly degrade the geometric and photometric quality of the mosaic. Existing solutions manage to deal with strong glares which corrupt the retinal content entirely while leaving aside the correction of semi-transparent specular highlights and lens flare. This introduces ghosting and information loss. Moreover, related generic methods do not produce satisfactory results in SLIM. Therefore, we propose a better alternative by designing a method based on a fast single-image technique to remove glares and the notion of the type of semi-transparent specular highlights and motion cues for intelligent correction of lens flare.

Finally, we solve the problem of automatic multi-modal registration of FA and SLIM. There exist a number of related works on multi-modal registration of various retinal image modalities. However, the majority of existing methods require a detection of feature points in both image modalities. This is a very difficult task for SLIM and FA. These methods do not account for the accurate registration in macula area - the priority landmark. Moreover, none has developed a fully automatic solution for SLIM and FA. In this thesis, we propose the first method that is able to register these two modalities without manual input by detecting retinal features only on one image and ensures an accurate registration in the macula area.

The description of the extensive experiments that were used to demonstrate the effectiveness of each of the proposed methods is also provided. Our results show that *(i)* using our new local refinement procedure for drift reduction significantly ameliorates the to drift reduction allowing us to achieve an improvement in precision over the current solution employed in the TrackScan; *(ii)* the

proposed methodology for correction of light-related artifacts exhibits a good efficiency, significantly outperforming related works in SLIM; and *(iii)* despite our solution for multi-modal registration builds on existing methods, with the various specific modifications made, it is fully automatic, effective and improves the baseline registration method currently used on the TrackScan.

Résumé

Cette thèse est issue d'un accord CIFRE entre le groupe de recherche EnCoV de l'Université Clermont Auvergne et la société Quantel Medical (www.quantel-medical.fr). Quantel Medical est une enterprise spécialisée dans le développement innovant des ultrasons et des produits laser en ophtalmologie. Cette thèse présente un travail de recherche visant à l'application du diagnostic assisté par ordinateur et du traitement des maladies de la rétine avec une utilisation du prototype industriel TrackScan développé par Quantel Medical. Plus précisément, elle contribue au problème du mosaicing précis de l'image de la lampe à fente (SLIM) et du recalage automatique et multimodal en utilisant les images SLIM avec l'angiographie par fluorescence (FA) pour aider à la photo coagulation pan-rétiienne naviguée. Nous abordons trois problèmes différents.

Le premier problème est lié à l'accumulation des erreurs du recalage en SLIM., il dérive de la mosaïque. Une approche commune pour obtenir la mosaïque consiste à calculer des transformations uniquement entre les images temporellement consécutives dans une séquence, puis à les combiner pour obtenir la transformation entre les vues non consécutives temporellement. Les nombreux algorithmes existants suivent cette approche. Malgré le faible coût de calcul et la simplicité de cette méthode, en raison de sa nature de 'chaînage', les erreurs d'alignement s'accumulent, ce qui entraîne une dérive des images dans la mosaïque. Nous proposons donc d'utilise les récents progrès réalisés dans les méthodes d'ajustement de faisceau et de présenter un cadre de réduction de la dérive spécialement conçu pour SLIM. Nous présentons aussi une nouvelle procédure de raffinement local.

Deuxièmement, nous abordons le problème induit par divers types d'artefacts communs á l'imagerie SLIM. Ceus-sont liés à la lumière utilisée, qui dégrade considérablement la qualité géométrique et photométrique de la mosaïque. Les solutions existantes permettent de faire face aux éblouissements forts qui corrompent entièrement le rendu de la rétine dans l'image tout en laissant de côté la correction des reflets spéculaires semi-transparents et reflets des lentilles. Cela introduit des images fantômes et des pertes d'information. En outre, les méthodes génériques ne produisent pas de résultats satisfaisants dans SLIM. Par conséquent, nous proposons une meilleure alternative en concevant une méthode basée sur une technique rapide en utilisant une seule image pour éliminer les éblouissements et la notion de feux spéculaires semi-transparents en utilisant les indicativons de mouvement pour la correction intelligente de reflet de lentille.

Finalement, nous résolvons le problème du recalage multimodal automatique avec SLIM. Il existe une quantité importante de travaux sur le recalage multimodal de diverses modalités d'image rétinienne. Cependant, la majorité des méthodes existantes nécessitent une détection de points clés dans les deux modalités d'image, ce qui est une tâche très difficile. Dans le cas de SLIM et FA ils ne tiennent pas compte du recalage précis dans la zone maculaire - le repère prioritaire. En outre, personne n'a développé une solution entièrement automatique pour SLIM et FA. Dans cette thèse, nous proposons la première méthode capable de recolu ces deux modalités sans une saisie manuelle, en détectant les repères anatomiques uniquement sur une seule image pour assurer un recalage précis dans la zone maculaire.

La description des expériences approfondies utilisées pour démontrer l'efficacité de chacune des méthodes proposées est également fournie. Nos résultats montrent que *(i)* notre nouvelle procédure de raffinement local pour la réduction de la dérive, qui peut également être appliquée dans d'autres champs tels que le suivi des objets dans le domaine non médical, contribue de manière significative à la réduction de la dérive, permettant d'atteindre une amélioration de la précision par rapport à la solution actuelle utilisée dans TrackScan; *(ii)* la méthodologie proposée pour la correction des artefacts liés à la lumière présente une bonne efficacité, surpassant significativement les travaux connexes dans SLIM; et *(iii)* malgré que notre solution pour le problème de recalage multimodal s'appuie fortement sur les méthodes existantes, avec les différentes modifications spécifiques apportées, elle est entièrement automatique, efficace et améliore considérablement la méthode de recalage de base actuellement utilisée dans TrackScan

Manuscript organization

This manuscript contains 8 chapters. The following list provides a brief summary of every chapter.

Chapter 1: Introduction serves as the basic introduction required to fully understand the medical context, terms and scientific objectives. It also provides a list of the author's publications.

Chapter 2: Background is dedicated to a description of necessary theoretical background on the subject of image registration and the related challenges commonly faced by the researchers. It aims to ensure the full comprehension of the basics related to the thesis objectives.

Chapter 3: Previous Work is dedicated to a comprehensive overview of the image registration problem and corresponding issues such as accumulated registration errors, illumination artifacts and multi-modal registration. It also provides a literature review of mosaicing methods in retinal imaging.

Chapter 4: Comparative Study of Transformation Models presents describes author's first contribution - a comparative study of various geometric transformation models applied to mosaicing of retinal images obtained with slit-lamp. It describes an efficient point correspondence based framework for transformation model evaluation in a typical closed loop motion scenario. It also introduces a new measure for accumulated drift.

Chapter 5: Drift Reduction describes author's second contribution - a novel approach to reduce accumulated registration drift. It introduces a new local refinement procedure and a new measure for accumulated drift.

Chapter 6: Handling Reflection Artifacts presents author's third contribution - an effective technique to detect and correct illumination artifacts of different degrees in SLIM. A two stage methodology is described along with validation results on patients presenting with healthy and unhealthy retina. It also introduces a new measure of global photometric image quality.

Chapter 7: Angio2SLIM: Automatic Multi-modal Registration describes an automatic multi-modal registration method called Angio2SLIM which automates the process of registering FA images and SLIM. The detailed validation on different publicly available mono-modal and multi-modal retinal image datasets is also included.

Chapter 8: Conclusion provides a summary of the manuscript and the conclusions derived from the results of the presented work. It also gives an insight on the research perspectives.

Contents

1 Introduction **1**
1.1 Inside the human eye: a wider view of the retina . . . 2
1.1.1 Basic structure of the eye . . . 2
1.1.2 Retinopathies and clinical diagnosis . . . 2
1.1.3 Standard treatment: pan-retinal photocoagulation . . . 4
1.1.4 Slit-lamp based NPRP with TrackScan . . . 6
1.2 Problem formulation and thesis objectives . . . 10
1.3 Summary of contributions . . . 10
1.4 List of publications . . . 11

2 Background **13**
2.1 Image registration . . . 14
2.1.1 Application . . . 14
2.1.2 Formulation and general pipeline . . . 14
2.1.3 Feature space and matching strategy . . . 16
2.1.4 Search space and search strategy . . . 17
2.1.5 Similarity metrics . . . 20
2.1.6 Warping and blending . . . 21
2.1.7 Classification of registration methods . . . 22
2.1.8 Factors complicating image registration . . . 24
2.1.9 Assessment of registration accuracy . . . 26
2.2 Light-related imaging artifacts . . . 26
2.3 Summary . . . 28

3 Previous Work **29**
3.1 Scope . . . 30
3.2 Image mosaicing . . . 30
3.2.1 Application to retinal imaging . . . 30
3.2.2 Assessment of transformation models . . . 36
3.2.3 Reduction of accumulated registration errors . . . 37
3.3 Multi-modal image registration . . . 39
3.3.1 Medical imaging applications . . . 39
3.3.2 Retinal image modalities . . . 41
3.4 Detection and correction of light-related imaging artifacts . . . 43
3.4.1 Specular highlight correction in medical imaging . . . 43
3.4.2 Specular highlight correction in the non-medical domain . . . 44

3.4.3 Application to retinal imaging . . . 45
3.5 Summary . . . 45

4 A Comparative Study of Transformation Models 49
4.1 Motivation . . . 50
4.2 Slit-lamp imaging and geometric assumptions . . . 50
4.3 Transformation models and evaluation framework . . . 50
4.3.1 Data acquisition . . . 51
4.3.2 Selection of pairwise point correspondences . . . 51
4.3.3 Transformation parameter estimation . . . 52
4.3.4 Evaluation . . . 54
4.4 Experimental results and discussion . . . 54
4.4.1 Part I: effect of model complexity . . . 55
4.4.2 Part II: effect of the number of points . . . 57
4.5 Conclusion . . . 58

5 Drift Reduction 59
5.1 Motivation . . . 60
5.2 Methodology . . . 60
5.2.1 Mosaicing initialization . . . 60
5.2.2 Motion estimation . . . 61
5.2.3 Prediction . . . 61
5.2.4 Track correction . . . 61
5.2.5 Key-frame instantiation and Local Bundle Adjustment . . . 62
5.3 Experimental results and discussion . . . 63
5.3.1 Dataset acquisition . . . 63
5.3.2 Evaluation . . . 63
5.4 Conclusion . . . 68

6 Handling Reflection Artifacts 69
6.1 Motivation . . . 70
6.2 Methodology . . . 71
6.2.1 Single-image glare removal and retina segmentation . . . 71
6.2.2 Multi-image lens flare correction: content-aware blending . . . 72
6.3 Experimental results and discussion . . . 74
6.3.1 Dataset and evaluation strategy . . . 74
6.3.2 Single-image glare removal and retina segmentation . . . 74
6.3.3 Multi-image highlight correction: content-aware blending . . . 77
6.4 Conclusion . . . 77

7 Angio2SLIM: Automatic Multimodal Registration 81
7.1 Motivation . . . 82
7.2 Angio2SLIM . . . 83
7.2.1 Retinal features detection in a reference FA image . . . 83
7.2.2 Automatic matching using SOM and LBP . . . 85
7.2.3 Non-rigid image registration with the normalized quadratic model . . . 89
7.3 Experimental results and discussion . . . 90
7.3.1 Datasets and ground truth . . . 90
7.3.2 Inclusion of the priority landmark detection . . . 92
7.3.3 Automatic point matching . . . 93

7.3.4 Multi-modal registration: comparative results 94
7.4 Conclusion 98

8 Conclusions **99**
8.1 Conclusion 100
8.2 Future work 101

Abbreviations **105**

List of Figures **109**

List of Tables **111**

Bibliography **123**

Chapter 1

Introduction

In this chapter we provide a detailed description of the reason that compels this thesis. In §1.1 we give a general information about the structure of the human eye and explain the medical context to ensure understanding of the necessary terminology which will be used in the following chapters. This includes the description of the retinal disorders, the tools used for their diagnosis, the treatment strategies applied and corresponding medical devices. We also present one of such medical devises, the industrial prototype which we work on within the scope of this thesis. This is followed by stating its limitations which then helps us to formulate our research problem and thesis objectives in §1.2. We also provide a summary of the contributions and related publications resulted from this thesis in §1.3 and §1.4 respectively.

Contents

1.1 Inside the human eye: a wider view of the retina 2

1.1.1 Basic structure of the eye 2

1.1.2 Retinopathies and clinical diagnosis 2

1.1.3 Standard treatment: pan-retinal photocoagulation 4

1.1.4 Slit-lamp based NPRP with TrackScan 6

1.2 Problem formulation and thesis objectives 10

1.3 Summary of contributions 10

1.4 List of publications 11

1.1 Inside the human eye: a wider view of the retina

Human vision is a very complex process that is not completely understood, despite hundreds of years of intense study and modeling. It gives our bodies the ability to perceive the surrounding environment and requires communication between its major sensory organ - the eye, and the brain - the core of the central nervous system, to interpret light waves as images. Our vision depends mainly on the eye which is one of the most complicated structures on earth. It requires many components to allow our advanced visual capabilities.

1.1.1 Basic structure of the eye

Figure 1.1 shows the cross section of the healthy human eye and illustrates the most important parts. The *sclera* is an outer membrane, which protects and supports the shape of the eye. It is what gives most of the eyeball its white color. The *cornea* is the transparent layer forming the front of the eye. The *choroid* layer provides nutrition to the eye and consists of blood vessels, iris, pupil and lens. The *iris* is the pigmented anterior portion of choroid which gives color to the eye. The *pupil* is a central opening of the iris, which controls the amount of light entering to the eye just as aperture controls the light coming to the image sensor in the modern photo cameras. The *lens* is made of concentric layers of fibrous cells where the image formation process begins. The *retina* is a photosensitive area composed of nerve cells that line the bottom of the eye and transform the luminous signal into an electrical signal. It is then sent to the visual areas of the brain to be interpreted. The *optic nerve* transfers information of the projected image from the retina to the brain. The head of the optic nerve, called *optic disk*, does not contain receptors itself, and is thus the blind spot of the eye. The *macula* is an oval spot near the center of the retina with a diameter of about 1.5 millimeters. Finally, the *fovea* is near the center of the macula and it contains packed cone cells. Due to high amount of light sensitive cells, the fovea is responsible for the most accurate vision.

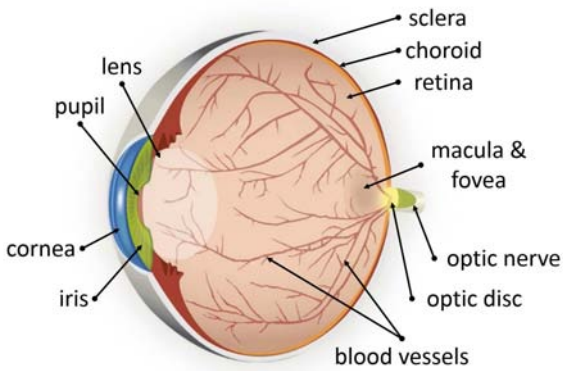


Figure 1.1: Important components of the human eye.

1.1.2 Retinopathies and clinical diagnosis

Global estimates on visual impairment reported by the World Health Organization (WHO) show that the principal cause of blindness is cataract - a clouding of the lens in the eye, for about 51% [Pascolini and Mariotti, 2011]. However, retina related disorders such as Diabetic Retinopathy (DR) and Age-related Macular Degeneration (AMD), both referred as retinopathies, are the leading causes of preventable blindness among working populations in economically-developed societies.

The DR is composed of a group of lesions found in the retina as a complication associated with diabetes among individuals suffering from the disease for several years. The abnormalities occur in predictable progression with minor variations in the order of their appearance. DR is considered to be the result of vascular changes in the retinal circulation. It is also known as Non-proliferative Diabetic Retinopathy (NPDR). It progresses into a Proliferative Diabetic Retinopathy (PDR) with the growth of new blood vessels referred to as *neovascularization*. Macular edema (the thickening of the central part of the retina) can significantly decrease visual acuity.

AMD is a condition affecting older people, which is described as loss of the person's central field of vision. It occurs when the macula develops degenerative lesions causing circulatory insufficiency with reduction in the blood flow to the macular area. Genetic factors as well as environment and lifestyle (*e.g.* smoking, hypertension, obesity, etc.) play a key role in the formation of the disease.

Classification As the incidence of the aforementioned types of retinopathies gradually increases, there is the possibility that more individuals will suffer from eye complications which, if not properly managed, may lead to permanent eye damage. Figure 1.2 illustrates the visual signs which can be observed in patients with various stages of DR and AMD.

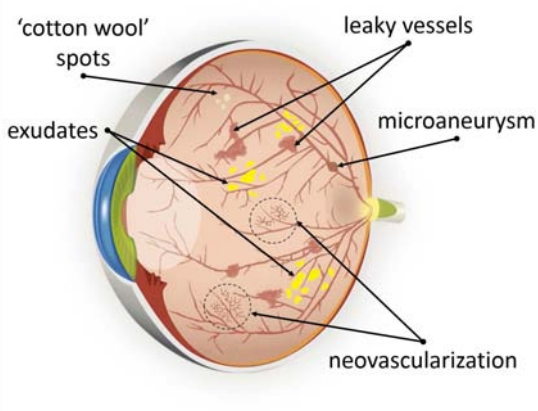


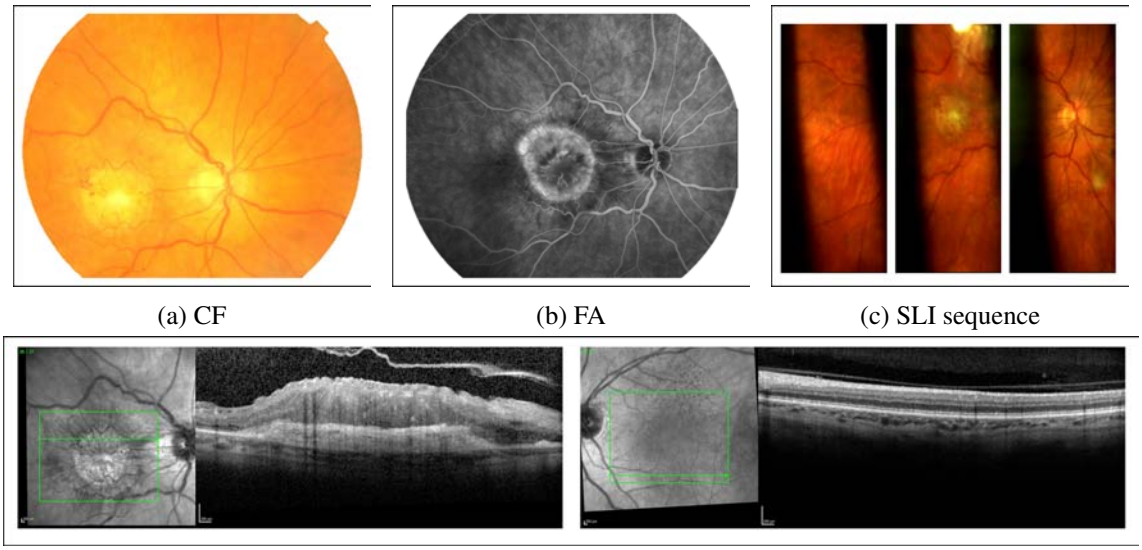
Figure 1.2: Visual signs of diabetic retinopathy and age-related macular degeneration.

The types of DR can be broadly classified as follows:

- **Mild/early stage.** Mild NPDR is considered as the first step in the evolution of the DR. *Microaneurysms* start to occur at this stage when the tiny blood vessels in the retina begin to swell. Early AMD is defined by the presence of numerous small or intermediate macular lesions.

- **Moderate stage.** Moderate NPDR is characterized by multiple microaneurysms and nerve fiber layer infarctions known as *cotton-wool spots*. Moderate AMD characterized by either extensive drusen of small or intermediate size, or any drusen of large size. Drusen are yellow deposits under the retina which are made up of lipids, a fatty protein.

- **Advanced/severe stage.** Severe NPDR show an increased number of microaneurysms and indicates intraretinal microvascular abnormalities. PDR is the most advanced form. It is characterized by the development of large areas of irreversible retinal ischemia. Ischemic retinal cells produce vascular growths that trigger the proliferation of abnormal neovascularization that are responsible for more severe complications: bleeding and detachment of the retina. Advanced AMD is defined by the presence of either macular atrophy or choroidal neovascular membrane. The presence of *exudates* define the exudative form of AMD, caused by the breakdown of the blood-retina barrier, allowing leakage of proteins.



(a) CF (b) FA (c) SLI sequence

(d) IR+OCT

Figure 1.3: Examples of retinal image modalities commonly used for diagnosis and treatment planning. Multimodal imaging performed on a patient's left eye with AMD. (d) shows a zoom-in on a region of IR with a green arrow indicating the corresponding OCT slices of the AMD affected eye (left) and healthy eye (right).

Diagnosis With special imaging and examination methods it is possible to perform direct *in vivo* non-invasive observation of the retina and identify the type and stage of the retinopthy. An ophthalmologist has several diagnostic imaging modalities at his disposal [Khaderi et al., 2011]. Examples of these pre-operative images are shown in Figure 1.3. Color Fundus Photography (CF) is obtained with a fundus camera (a low power microscope with an attached camera) which photographs the interior surface of the eye, including the retina, retinal vasculature, optic disc and macula. FA is acquired using a fluorescent dye and a specialized angiographic camera to examine the circulation of the retina and choroid. Slit-Lamp Image (SLI) is obtained by a biomicroscope coupled with a slit-lamp during ophthalmoscopy. Optical Coherence Tomography (OCT) benefits from light to capture micrometer-resolution, three-dimensional images of the eye's anterior segment and retina. Finally, Infra-red (IR) fundus images, which are shown in Figure 1.3d as a square region attached to the OCT image, are obtained by a scanning laser ophthalmoscope and only infra-red wavelengths is utilized.

1.1.3 Standard treatment: pan-retinal photocoagulation

For more than 50 years, laser technology has evolved to become an essential tool in the treatment of diabetic retinopathies. Laser Pan-retinal Photocoagulation (PRP) is considered the standard treatment worldwide. Two key studies of PRP (by Diabetic Retinopathy Study Research Group and Early Treatment Diabetic Retinopathy Study Research Group) reporting clinical trials have shown that laser therapy reduces the risk of severe vision loss by at least 50% among patients with PDR [DRS, 1981, ETDRS, 1991]. PRP is performed using the coagulating effect of a laser beam directed to the retina via contact lens of strong convergence to destroy the pathological zones between the macula and the periphery. A simplified scheme of the process is shown in Figure 1.4. The beam's radius is focused precisely on the affected area and its wavelength is chosen in relation to the absorption levels. When the energy from a strong light source is absorbed by the retinal pigment epithelium and is converted into thermal energy, coagulation necrosis occurs. Thermal burns help to stabilize the progress of the disorder. In addition, selective laser photocoagulation in the macular area is also an effective treatment

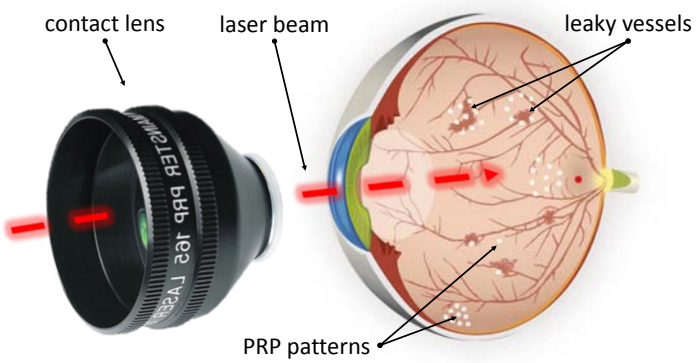


Figure 1.4: Simplified general illustration of pan-retinal photocoagulation with a contact lens.

for macular edema. Patients require anesthesia for the procedure. Most patients undergo PRP under topical anesthesia such as specific eye drops while others may require various injections of lidocaine or even general anesthesia (used for infants, children, and patients with problems in compliance).

Laser delivery The aforementioned clinical trials established PRP as the standard treatment of choice for complications of DR for many decades. Even though, clinically effective, retinal laser photocoagulation leads to significant side effects including reduced night vision, decrease in central and peripheral vision and disruption of the retinal anatomy through scarring. In search for solutions to safe retinal tissue yet achieving desired therapeutic effect, laser technology has evolved [Kozak and Luttrull, 2015]. A control over laser power, spot size and precision was achieved along with the creation of micropulsed lasers allowing for a selective treatment. The variety of laser treatment patterns as the one shown in Figure 1.4 now allow one to minimize the collateral damage. While the aforementioned attempts were focused on laser adjustments, the major developments took place on side of laser delivery.

Conventional slit-lamp based systems, which date back to the 1980s, are the common technology that every ophthalmologist is familiar with. The PRP is performed via slit-lamp biomicroscope and a contact lens. The laser is attached to the typical ophthalmic slit-lamp devise used for biomicroscopical examination of the retina and the laser energy is delivered in a coaxial fashion. The patient is placed in a seated position, and the chin placed on the chin-rest as shown in Figure 1.5. A contact lens, which focuses the laser onto the retina, is placed against the cornea with clear coupling agent. Typically a wide angle or mirrored lens is used. The laser is then fired transcorneally through this contact lens, focused on the retina. The laser delivery is manually controlled by an ophthalmologist and consists of long pulse durations, typically 100 milliseconds, large single-spot size (200-500μ), with 200-250mW of power applied.

The first attempts to make photocoagulation a completely automated procedure involved image recognition software and eye tracking [Wright et al., 2000]. However, the complexity of such systems prevented their commercial introduction and acceptance in clinical practice back in the 2000s. A semi-automatic pattern scanning photocoagulator (PASCAL, Topcon Medical Laser Systems Inc) was introduced by OptiMedica Corp. in 2005 [Blumenkranz et al., 2006]. It delivers a pattern of multiple burns in the same or shorten amount of time that conventional lasers take to deliver one burn. The speed of delivery allows newer lasers to reduce the pulse duration to 10-30 milliseconds per spot, which is balanced by many more total spots. Short duration lasers provide patients with more comfort than long-duration PRP does. The PASCAL system is fully integrated with a touch screen Graphical User Interface (GUI), however it remains the same in terms of imaging and illumination (slit-lamp

optics), slit-lamp mounted micromanipulator and spot positioning. Ergonomic features were added for the physician and patient's comfort [Blumenkranz et al., 2006].

The second major development was the introduction of a new fully automatic laser platform called NAVILAS (OD-OS, IncĠermany) guided by diagnostic imaging and stabilized using eye tracking [Kozak et al., 2011, Chalam et al., 2012]. Apart from offering retina navigation, it has similar technical specifications as PASCAL (single or predetermined pattern array, 10-30ms pulse duration). NAVILAS integrates live color fundus imaging, infra-red imaging, and fluorescein angiography with a photocoagulator system. It can be short or long-duration and uses either single-spot or pattern-spot arrays that reach all the way to the peripheral retina. This system includes retinal image acquisition, annotation of the images to create a detailed treatment plan, and then automated delivery of the laser to the retina according to the treatment plan. The physician controls laser application and the systems assist with prepositioning the laser beam.

Each means of laser delivery has both merits and challenges. Compared with conventional slit-lamp laser delivery, both PASCAL and ANVILAS use shorter laser pulses, cause relatively less thermal damage to adjacent retinal tissues and can therefore produce relatively fever side effects [Kozak et al., 2011, Chhablani et al., 2014, Inan et al., 2016, Stewart, 2017]. Even if the visible area of the retina is smaller with the slit-lamp laser delivery compared to the image modalities integrated in NAVILAS, it has the advantage of more precision, magnification and control. Additionally, contact lenses can offer some stabilization of wandering eye movements and stabilize the lids for those prone to muscle contractions around the eye.

1.1.4 Slit-lamp based NPRP with TrackScan

While the fundus camera based system [Chalam et al., 2012] is considered the best NPRP system, the magnification and control offered by the slit-lamp still makes it a very popular choice in the clinical environment [Asmuth et al., 2001]. In this context the development of a platform to combine conventional slit-lamp laser delivery with computer-assisted navigation is on demand. This has a benefit to assist ophthalmologists in their preoperative planning, intraoperative navigation and photographic documentation while preserving the familiarity with conventional slit-lamp biomicroscopy.

TrackScan platform and SLIM Recently, a computer assisted slit-lamp based industrial prototype has been developed in QuantelMedical. The prototype combines real-time High Definition (HD) imaging, pre-operative planning and intra-operative navigation. It also provides the basic functionality for multi-modal registration of diagnostic images. Figure 1.5 provides an illustration of the platform. The imaging setup is based on the eyepiece and microscope optics of the slit-lamp and the magnifying contact lens attached to the eye such that slit illumination is projected onto the retina. This setup is used to perform retinal examination and treatment where the ophthalmologist typically explores the retina in a closed-loop manner starting from the optic disc. The platform is composed of:

- a slit-lamp coupled with a biomicroscope;
- two HD sensors which capture 60 images per second. Each camera corresponds to each side of the binocular microscope;
- a computer equipped with an image acquisition board and a Graphics Processing Unit (GPU) powered graphics card. The acquisition board embeds an Field-programmable Gate Array (FPGA) integrated circuit and memory to delegate calculations on incoming video streams;
- a complete laser system (laser source + scanner + zoom), the processing part of which is integrated into the slit-lamp and the source is interfaced with the computer;

The slit-lamp biomicroscope makes it possible to obtain a SLI with a high resolution which allows them to perform an accurate laser shot. The counterpart is that the visualization is local and

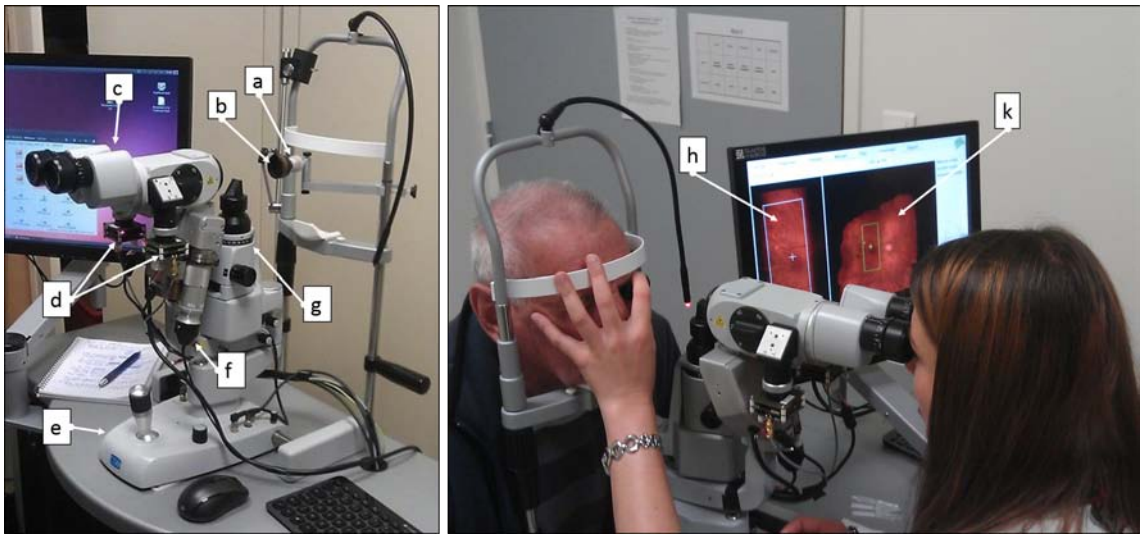
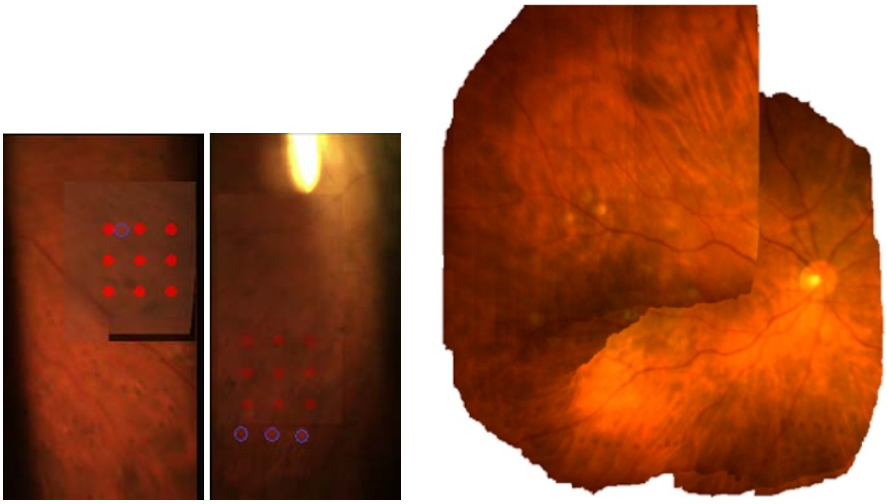


Figure 1.5: The main components of the TrackScan platform developed in QuantelMedical and an illustration of SLIM during retinal examination of a patient at University Hospital of Saint-Étienne, France. (a) phantom eye; (b) contact lens; (c) binocular microscope; (d) HD sensors; (e) moving base; (f) laser supply; (g) slit-lamp; (h) live SLI sequence; (k) intra-operative retina map.

does not help the practitioner in his therapeutic act. Computer-assisted retina mosaicing method for view expansion using a slit-lamp device proposed in [Richa et al., 2014] has been integrated into the TrackScan, allowing view expansion by building a map of the retina in real time. From now on and further we define this retinal map construction as SLIM and use this abbreviation to refer to the retinal mosaic obtained with SLIM as a new retinal image modality which can be used for diagnostic purposes.

Limitations Although the method proposed by [Richa et al., 2014] is capable of producing mosaics with a very good definition of the retina, while being robust to the variability of the patients, it has a number of problems. In addition, the multi-modal registration module does not provide the desired outcome. These limitations can be summarized as follows:

Mosaicing Drift. A simulation of the localization of the visualized area and the laser delivery on the treatment plan with multiple points has been conducted to verify if accurate navigation could be achieved. The first results demonstrated that the target area was the area provided in the treatment plan. Nevertheless, the positioning error of each laser spot was still significant. This is illustrated in Figure 1.6a. This is due to the geometrical errors accumulated during the creation of the mosaic and the localization errors due to the movements of the eye. Moreover, lens distortions and tracking errors contribute to the mosaicing drift and therefore degrade the quality of retinal mapping. The misalignment of vessels caused by inaccurate mapping can be seen in Figure 1.6b. These errors are caused by the transformation model used in the mapping process which does not take into account the deformations created by the combination of the wide angle lens and the lens of the microscope.



(a) (b)

Figure 1.6: TrackScan limitations: uncorrected mosaicing drift. (a) a mismatch of treatment plan (red) with the laser spots (blue); (b) a visually distinctive vessel misalignment.

Specular reflections. The reflections of the light source on the lens of strong convergence or the cornea of the patient create specular reflections that are difficult to separate from the retina. The movement of these artifacts can be different from the apparent movement of the retina, which often leads to degradation or loss of follow-up. Specular highlight removal as implemented in the mosaicing method of [Richa et al., 2014] has a limited performance under practical conditions as the light intensity and the gain of the camera vary significantly between the patients. The uncorrected specular highlights as they occur during slit-lamp examination among patient is shown in Figure 1.7.

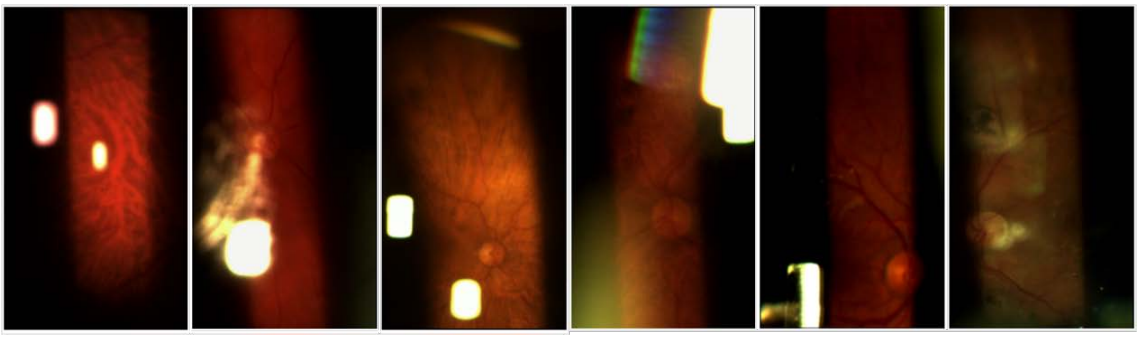


Figure 1.7: TrackScan limitations: uncorrected illumination artifacts of different degrees.

Multi-modal registration. A semi-automatic solution implemented on the TrackScan requires an operator to manually select corresponding points between the final retinal mosaic and the angiography image. Figure 1.8 shows an example of a pair of SLIM and FA images and the result of the registration. The preliminary results revealed heavy distortion of the images and it is not acceptable for the use in the treatment planning and intra-operative navigation as assessed by the experts. Moreover, a more complex transformation model shall be used in the process of registration as opposed to the currently implemented rigid model that is capable of recovering the translations and rotations only.

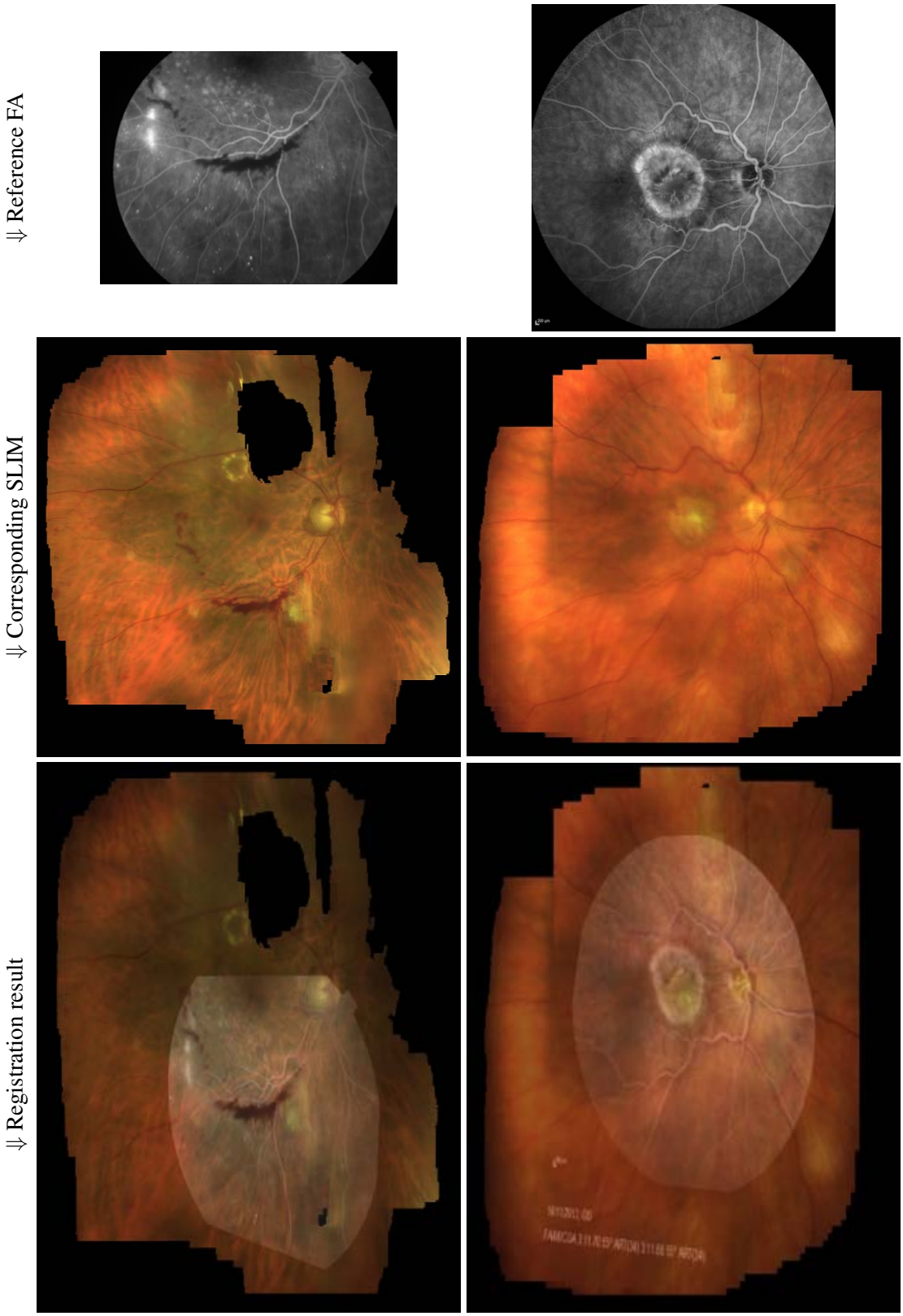


Figure 1.8: TrackScan limitations: multi-modal registration.

1.2 Problem formulation and thesis objectives

The aforementioned limitations of TrackScan correspond to the problems we address in this thesis. The focus is to improve its functionality. Specifically, we formulate our objectives as a triplet and provide the description of the actions we set to achieve our goals as follows:

Objective 1 - reduction of the mosaicing drift to improve the mapping precision

* Assess the effects of various transformation models' complexity and the number of point correspondences on the accumulation of registration errors in SLIM.
* Review existing works on the problem of accumulated drift in mosaicing applications with emphasis on long image sequences obtained with closed loop motion.
* Propose methods and techniques for drift reduction dedicated to the case of SLIM.
* Evaluate the proposed solution and compare it to the baseline method [Richa et al., 2014].

Objective 2 - enhancement of the global photometric quality by minimizing the illumination artifacts

* Review existing works on the problem of illumination artifacts in generic application, medical imaging applications with emphasis on retinal imaging.
* Identify specific types of specular highlights which can be detected and corrected and design a dedicated solution for SLIM.
* Assess the proposed solution with quantitative and qualitative evaluation.

Objective 3 - automation of the process of multi-modal registration and improvement of the registration accuracy

* Review existing literature on multi-modal registration with emphasis on retinal imaging.
* Develop strategies and techniques to minimize user intervention.
* Propose an automatic refinement of the registration to improve accuracy.
* Quantitatively evaluate the proposed solution.

1.3 Summary of contributions

The research work completed within the scope of this dissertation was driven by the application of image mosaicing and image blending in the challenging environment of the long slit-lamp retinal image sequences. We demonstrate the significant improvement over the existing mosaicing method implemented on the slit-lamp based n NPRP industrial prototype. A summary of our contributions is given below and they are detailed in chapters 3, 4, 5 and 6 where the conclusions made in one chapter defines the basis for the approach presented in the next chapter, *i.e.* one leads to the other.

1. We propose a new evaluation framework and error metric to assess the amount of accumulated mosaicing drift. Despite the variety of works which report on different transformation models for retinal image registration, only a few address their comparison and evaluation. These works, however, do not consider the mosaicing of long image sequences obtained in a closed- loop motion, which is typical in examination with the slit-lamp as we have shown in [Prokopetc and Bartoli, 2016a]. In addition, we derived a new mathematical normalization procedure for the quadratic transformation model which helps to improve the model fitting.

2. The fairness of the conclusions on the choice of the right transformation model to serve our mosaicing application [Prokopetc and Bartoli, 2016a] led us to proceed with drift reduction. As our main contribution to the problem of video sequence mosaicing we combined several existing techniques to reduce accumulated mosaicing drift which resulted in a novel approach that we presented in [Prokopetc and Bartoli, 2016b]. We also derived a new local refinement procedure and a new measure for accumulated drift.

3. Our contribution to the problem of various types of illumination artifacts is a SLIM dedicated method. Glare eliminates all information in the affected pixels, and the other types of reflections can introduce artifacts in feature extraction algorithms, which are critical in our application. As opposed to existing works, we effectively combine standard techniques to address the problem of strong glares, lens flare and haze. We also introduced a novel metric for global photometric quality which we presented in [Prokopetc and Bartoli, 2017a] and [Prokopetc and Bartoli, 2017b].

4. A first fully automatic solution to the specific case of muti-modal retinal registration of FA and SLIM is our final contribution. We applied an existing neural network based technique to establish point correspondences and introduced a novel data driven measure of correspondence quality. The results provide an improvement over the method used in TrackScan. This work, however, can be extended and is currently in progress. It is planned to be submitted for publication in the journal of Medical Image Analysis.

1.4 List of publications

This thesis is a monograph, which contains original material. It is largely based on the publications mentioned in the following list.

[Prokopetc and Bartoli, 2016a]
K. Prokopetc and A. Bartoli, "A comparative study of transformation models for the sequential mosaicing of long retinal sequences of slit-lamp images obtained in a closed-loop motion", *Int. J. Computer Assisted Radiology and Surgery* , 11(12): 2163-2172, (CARS, Heidelberg, Germany)

[Prokopetc and Bartoli, 2016b]
K. Prokopetc and A. Bartoli, "Reducing Drift in Mosaicing Slit-Lamp Retinal Images", *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), International Workshop on Biomedical Image Registration (WBIR)*, 533-540, (CVPRW (WBIR) 2016, Las Vegas, NV, USA)

[Prokopetc and Bartoli, 2017a]
K. Prokopetc and A. Bartoli, "SLIM: slit-lamp image mosaicing", *Congrés Francophone des Jeunes Chercheurs en Vision par Ordinateur*, (ORASIS 2017, Colleville-sur-Mer, France)

[Prokopetc and Bartoli, 2017b]
K. Prokopetc and A. Bartoli, "SLIM (Slit-Lamp Image Mosaicing): handling reflection artifacts", *Int. J. Computer Assisted Radiology and Surgery*(special issue for the 9th International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), 12(6): 911-920 (IJCARS (IPCAI) 2017, Barcelona, Spain)

Chapter 2

Background

This chapter aims to give the reader full comprehension of the basics related to our thesis objectives. We discuss the applications of image registration, explain the general pipeline and give insights on its every step. We also reflect on a classification of the image registration methods that can be found in the literature. We first give a general introduction to image registration in §2.1. Because our two main contributions are related to image mosaicing and multi-modal image registration, we also cover these special cases along with a problem of accumulated registration errors and other challenges. In addition we briefly discuss the evaluation approaches for the image registration algorithms. In §2.2 we address the problem of light related imaging artifacts, specifically different lens flare, glare and specular highlights. This is necessary for understanding the challenging nature of SLIM.

Contents

2.1 Image registration **14**
- 2.1.1 Application 14
- 2.1.2 Formulation and general pipeline 14
- 2.1.3 Feature space and matching strategy 16
- 2.1.4 Search space and search strategy 17
- 2.1.5 Similarity metrics 20
- 2.1.6 Warping and blending 21
- 2.1.7 Classification of registration methods 22
- 2.1.8 Factors complicating image registration 24
- 2.1.9 Assessment of registration accuracy 26

2.2 Light-related imaging artifacts **26**

2.3 Summary **28**

2.1 Image registration

Image registration is one of the most fundamental problems in computer vision and medical image analysis. It aims at finding an optimal image transformation to align two or more images of the same scene into a single integrated representation. This is often a crucial step in many image analysis tasks where the final information is gained from the combination of various data sources. It is widely used in fields such as medical imaging, remote sensing, robotics, astrophotography, quality control, to name a few. To get a picture of the scope, the comprehensive surveys on image registration methods and their applications published by [Brown, 1992, Zitova and Flusser, 2003] are good references. [Maintz and Viergever, 1998] and [Mani and Rivazhagan, 2013] provide dedicated surveys to medical image registration methods.

2.1.1 Application

There exist multiple ways to categorize the applications of image registration. Generally, they can be divided into four distinctive groups with respect to the image acquisition mode. A brief description is given bellow:

Multi-view registration For images of the same scene taken from different viewpoints, the aim of registration is to gain a larger 2D view. This is referred to as image mosaicing and allows one to construct mosaics of scenes which are generally very large to be captured using a single image. This is applied in remote sensing domain, for mosaicing of images of the surveyed area, in computer vision for object shape recovery, or in medical applications like SLIM.

Temporal registration When images are taken at different times, image registration allows one to evaluate the changes in the scene that appeared between two different image acquisitions. For instance, in remote sensing, this helps to monitor global land usage and landscape planning. It is widely used in visual surveillance for automatic detection of changes in a surveyed area. In the medical imaging domain, registration allows one to monitor the healing therapy or the progress evaluation of a disease.

Multi-modal registration When different sensors are used for imaging the same scene, registration provides the means for integration of information from different sources, modalities, and thus makes it possible to obtain a more complex and detailed scene representation. In remote sensing, multi-spectral satellite images can be registered together. Medical imaging takes advantage of multi-modal registration to gain information from sensors recording anatomical body structure with information from sensors monitoring functional and metabolic body activities, like registering Magnetic Resonance Image (MRI) with Positron Emission Tomography (PET).

Scene-to-model/Model-to-scene registration Images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance maps or another scene with similar content (another patient), 'average' specimen, etc. The aim is to localize the acquired image in the scene/model and/or to compare them. This includes registration of aerial or satellite data into maps, target template matching with real-time images, automatic quality inspection, comparison of the patient's image with digital anatomical atlases.

2.1.2 Formulation and general pipeline

Within the wide spectrum of image registration applications, this thesis focuses on techniques relevant to the automatic multi-view and inter-modal registration of regular 2D image data in retinal imaging. These problems can be treated an viriety of different ways. However, regardless the length of the

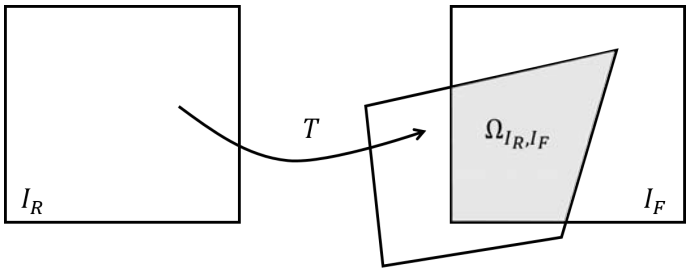


Figure 2.1: General principle of 2D image registration.

image sequence or image modality and for the sake of simplicity, the basic image registration process can be narrowed down to the case of registering two images. This is posed as an optimization problem.

Given a *reference* image I_R and the *floating* image I_F, the task of image registration consists in finding an optimal *transformation* T that maps from coordinates in the reference image to coordinates in the floating image, such that the transformed image $T(I_F)$ is *similar* to I_R. This is formulated as $T : \mathbb{R}^D \mapsto \mathbb{R}^D$, where $\mathbb{R}^D$ represents a D-dimensional data space with $D = 2$. This is achieved by determining the optimal parameters $\tilde{\theta}$ of the transformation T from the data through regularization and the similarity is defined through a *similarity measure* C. This can be formulated as:

$$\tilde{\theta} = \underset{\theta}{\operatorname{argmax}}\, C(I_R, T(I_F, \theta)) \tag{2.1}$$

where I_R is the image that is kept unchanged and used as a reference for registration. I_F is the image that is spatially warped to align with the reference image. $T(\cdot)$ is a space of allowed transformations to *warp* the floating image onto the reference image. C is the metric used to quantify the registration success and $\tilde{\theta}$ is an optimal set of parameters that makes C to reach its maximum. The C, which is used as an objective function in the optimization process, is a key aspect of image registration. The general principle of 2D image registration is illustrated in Figure 2.1 where Ω_{I_R,I_F} represents the pixels in the overlapping area of the two images which may vary with each estimate of C.

Due to the diversity of images to be registered and their specific nature it is impossible to design a universal method applicable to all registration tasks. Every method should take into account not only the assumed type of transformation but also imaging artifacts, required registration accuracy and application-dependent data characteristics. A general image registration pipeline, however, consists of the combination of several important elements common to all the methods. According to [Brown, 1992], every registration algorithm consists of 4 main components. A choice for each of the components can be made from a variety of alternatives. Below we list all of them and augment this list by adding *matching strategy*, *warping strategy* and *blending strategy* as these are also important choice that one has to make in the registration pipeline, either for mosaicing or pairwise mono-modal and multi-modal registration. Note that only basic concepts are presented here. For a more comprehensive overview of the components and the related methods we refer the reader to [Brown, 1992, Zitova and Flusser, 2003].

1. **Feature space** represents the distinctive information common to both images that will be used to establish image correspondences. Depending on the nature of the photographed scene and the registration approach the feature space may consist of image corners, pixels that belong to object edges, blobs/regions of interest points, ridges or an entire set of image pixels.

2. **Matching strategy** aims at finding the best correspondence in another image from the set of features. Depending on the types of features this can either be performed via feature description when a unique feature signature is computed in one image and compared to the feature descriptors obtained from another image; or a pixel-to-pixel correspondences established directly.

3. **Search space** is a space of transformations that are capable of aligning two given images.

4. **Search strategy** defines how to choose the next transformation from the search space, to be tested in the search for the optimal transformation.

5. **Similarity metric** determines the relative merit for each test in the search for an appropriate transformation. Search continues according to the search strategy until a transformation is found whose similarity criterion is satisfied.

6. **Warping strategy** defines how the input image is mapped into the output image using spatial transformation T to produce a *warp* ready for blending.

7. **Blending strategy** requires combining the colors of corresponding pixels in the overlap area of the images. It is often the case that significant global photometric differences can occur between images. If not corrected, this can give rise to unsightly seams in final registration result. The image blending function can be chosen to ameliorate this effect.

It is evident that each of the aforementioned components contributes significantly to the final result. The type of variation that is expected between the images determines the selection of individual components. Furthermore, the components are not independent and selection of one directly affects the choice for the others. Overall, the development of an image registration method is a complex problem of different interrelated components that have to be designed carefully to form a system that will give the best results. We provide more details on every component of the general registration pipeline in the following subsections.

2.1.3 Feature space and matching strategy

As mentioned above, the feature space is highly dependent on the image data. In medical imaging, extracted features often correspond to the location of landmarks which can be either natural (*i.e.* anatomical) or artificial (*i.e.* created intra-operatively or placed *post-mortum*). The latter is often the case if a *phantom* is used (*i.e.* an artificial replicate of an object). Generally features are *keypoints* in the image that have distinctive nature or all pixels can be considered as such. There exists many keypoint detectors which are used in computer vision applications. The basis for many detectors is the corner [Beaudet, 1978, Harris and Stephens, 1988, Lowe, 2004, Mikolajczyk and Schmid, 2004]. This builds on the assumption that the image gradient around a corner has at least two dominant directions. Besides, corners are repeatable and distinctive. Thus, corner detectors look for intersection points between two or more edge segments which is indicated by intensity changes in all directions. They either work directly with the brightness value of the images or extract object boundaries first then analyze its shape. Instead of using corners, local extrema of image intensity can serve as anchor points as well [Tuytelaars and Van Gool, 2004, Matas et al., 2004]. These points cannot be localized as accurately as corner points, since the local extrema in intensity are often rather smooth. However, they can withstand any monotonic intensity transformation and they are less likely to lie close to the border of an object resulting in a non-planar region. This last property is a major drawback when working with corner points. A saliency of a local region based on the contrast analysis can be exploited too [Kadir and Brady, 2001]. In the dense sampling approach no real keypoints are determined, but a dense grid of sample points is taken instead. This is useful for tracking applications. The main requirements to feature detectors can be summarized into a triplet: *1)* they need to be frequently spread over the image and easily and robustly detectable (*i.e.* repeatable detection); *2)* precise localization, meaning they need to have enough common elements even when the overlap of imaged scenes is small, or when object occlusions occur. Furthermore, the features need to have accurate localization property and should be immune to expected image variations; and *3)* distinctive content

(*i.e.* highly informative). Often in practice multiple types of features are extracted and combined to form *feature vectors*.

The aforementioned detectors precisely localize points with high repeatability. In order to compare these points *feature descriptors* over a region centered at the point should be computed. This provides a distinctive signature to every feature. Feature descriptors are usually designed as a function on the region, which is *scale invariant* (*i.e* the same for corresponding regions, even if they are at different scales), rotation and lighting. Average intensity, for example, is the same for corresponding regions even if they differ in size. There exist numerous generic image feature descriptors whichare widely used in many applications such as Scale Invariant Feature Transform (SIFT) [Lowe, 2004], Speeded-Up Robust Features (SURF) [Bay et al., 2006], Local Binary Patterns (LBP) [Ojala et al., 2002], Binary robust independent elementary features (BRIEF) [Calonder et al., 2012] and many others including their extensions and variations. Applications specific descriptors can also be a reasonable choice, *e.g.* Partial Intensity Invariant Feature Descriptor (PIIFD) [Chen et al., 2010] and its improved version [Ghassabi et al., 2013] are used in retinal imaging. An example of detected keyponts with corresponding SIFT descriptors are shown in Figure 2.2.

Figure 2.2: Keypoints detected on a photograph of the Cathedral of Clermont Ferrand. SIFT is used to describe the detected keypoints. Yellow circles visualize the position of the keypoint, the scale and orientation of the corresponding descriptor.

The matching strategy is chosen accordingly. The simplest matching approach is to match all feature vectors of one image to all of the other image, where the smallest Euclidean distance d is assumed to be the correct match. This often returns many false positives. A better strategy is to look for the second Nearest Neighbor after sorting the result with respect to the distance. The main assumption is that the best match should be significantly better than the second best match. As a special case, for matching high dimensional features, two algorithms have been found to be the most efficient: the randomized k-d forest and Fast Library for Approximate Nearest Neighbors (FLANN) [Muja and Lowe, 2009]. While the aforementioned methods are not suitable for binary representations with LBP and BRIEF, random forestrandom ferns classifiers [Ho, 1995] or vocabulary trees are used instead [Nister and Stewenius, 2006].

2.1.4 Search space and search strategy

Almost all image acquisition techniques involve unwanted geometric transformations. They are caused by perspective distortions, optical distortions due to lens errors or aberration, limitations of the acquisition process and so forth. Geometric transformations permit to eliminate, to a large extent, these distortions. Only after correcting these errors it would be possible to derive accurate metric measurements from the images and compare the same or similar objects in different images. Geometric transformation consists of two basic steps: *1)* determining the pixel coordinates in the transformed

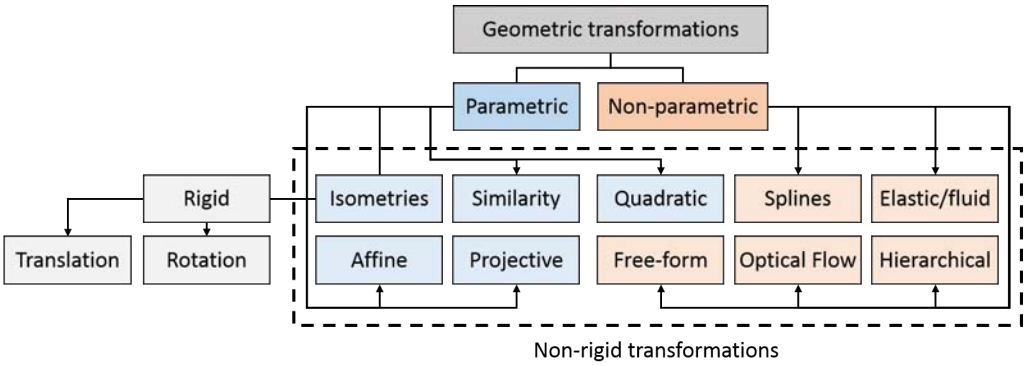


Figure 2.3: Different types of commonly used geometric transformations and their hierarchy.

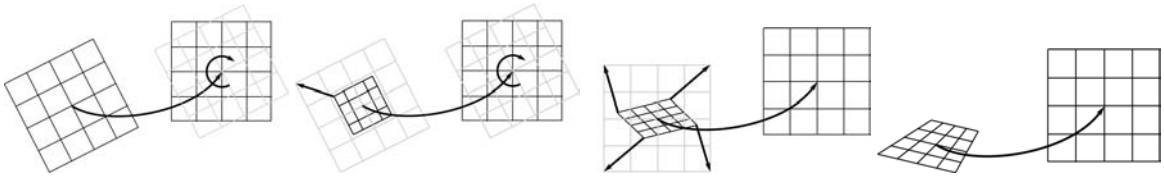
image and mapping of the coordinates (x, y) in the input image to the point (x', y') in the output image; *2)* determining the point which matches the transformed point and determining its brightness/color which is usually computed as an interpolation of several points in its neighborhood. The output coordinates generally do not fall onto exact pixel coordinates. The transformation T is either known in advance or can be determined from several known pixel correspondences in an original and transformed image pair.

Depending on the geometrical distortion one has to select the most appropriate model from a class of transformations. They can be broadly categorized to *parametric* and *non-parametric* as shown in Figure 2.3. If the transformation model has small and not varying number of parameters, the transformation is considered parametric. Otherwise, the transformation is called non-parametric (*e.g.* a set of parameters for each pixel of the image). Transformations are usually described by their complexity = Degrees of Freedom (DOF), which is the number of independent ways that the transformation can be changed. In general, increasing the number of DoFs allows the transformations greater scope.

Rigid and Parametric Non-rigid transformations An *isometry* (from *iso* = same, *metric* = measure), also know as rigid transformation, is a parametric transformation that preserves Euclidean distances (*i.e.* does not change the distance between any two points), as shown in Figure 2.4a. It is the simplest one that involves translations, rotations and their combination. In 2D space it is defined by 3 DoF. It is only appropriate for mono-modal registration of approximately rigid structures or as an initialization step for non-rigid registration methods.

A class of parametric non-rigid transformations builds on the rigid basis by adding additional complexity. Examples are shown in Figure 2.4. Thus, a first model that can be considered a non-rigid is a *similarity* model that preserves the shape by adding the uniform scale and defined by 4 DoF. An *Affine* transformation preserves points, straight lines, planes and parallelism. It has 6 DoF by augmenting the similarity model with shear anisotropy. A *Projective* transformation, also called *homography*, maps lines to lines (but does not necessarily preserve parallelism) has 8 DoF. All these models represent a planar map projection. The *Quadratic* transformation model is a simple extension of the linear affine model whose coefficients determine 12 DoF. It is often used in retinal imaging as will be shown in the next chapter. This transformation is generally referred to as *curved* nonlinear mapping which is based on 2nd-order polynomials and can be considered a special case of curved parametric model. This is because in general, curved transformations using polynomials of varying degree do not satisfy the condition of 'small and not varying number of parameters' mentioned above (*i.e.* the number of parameters depends on the degree of the polynomial).

Non-parametric Non-rigid transformations The registration techniques using non-parametric transformations are based on the assumption that a sparse set of corresponding points (control points) can



(a) rigid (b) similarity (c) affine (d) projective

Figure 2.4: Rigid and parametric non-rigid transformations.

be identified in the reference and floating image. The transformation then interpolates a smooth and dense deformation field between these control points. The general principle is illustrated in Figure 2.5.

The *Splines* transformations use a linear combination of radial bases functions instead of polynomials, *e.g.* the Thin-Plate Spline (TPS). The basic idea of TPS is to identify a sparse set of points in the reference image, search the corresponding set of points in the floating image, find a spline transformation, which interpolates the deformation field exactly at these points and smoothly varies between them. If the control points are not uniformly spaced, large errors may be obtained away from the control points. More complex transformations, such as *Elastic/fluid*, in which the deformation is controlled by a physical model that has taken into account the material properties, such as tissue elasticity or fluid flow, or the *Free-form*, in which any deformation is allowed, are also often used in medical imaging applications. The most general case of geometric transformation is the free-form transformation. These transformations can correct nonlinear distortions. Non-parametric transformations are no longer parametrized which gives an ill-posed problem. Thus, they are regularized by adding a penalty. Penalization, however, reduce the allowed deformations, but this is intended in most cases. Another non-parametric non-rigid method of transforming one image to another is to solve Partial Differential Equations (PDE). PDE based registration determines a pixel-to-pixel correspondence by computing a velocity field describing the apparent motion depicted between images. This is often referred to as *Optical flow* due to the work of [Horn and Schunck, 1981]. The *Hierarchical* non-parametric registration is yet another quite common non-rigid registration approach. The main idea is to apply a particular transformation while dividing the image to sub-images. Thus, at different levels of sub-division a higher complexity transformation can be used. The hierarchical approach is computationally very efficient as it carries over registration information from the previous levels. It is, however, prone to generate unwanted discontinuities.

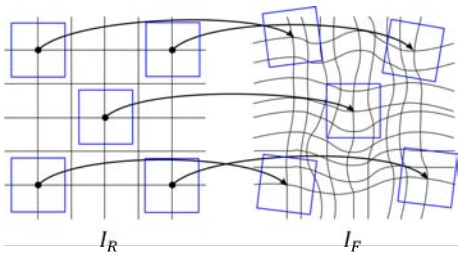


Figure 2.5: General principle of non-parametric non-rigid transformation.

The transformation model should be chosen according to the *a-priori* known information about the acquisition process and expected image variations as well. If no *a-priori* information is available the model should be general enough to handle all possible variations that might occur.

Optimization While considering the available computational resources, the search strategy has to yield a robust solution which is as close as possible to the optimal one. Thus, if considering an interative optimization, the next transformation from the search space can be obtained either in the

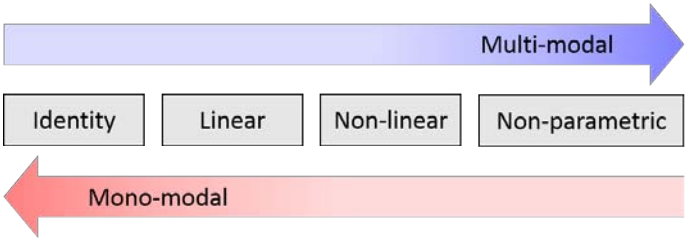


Figure 2.6: Different types of the pixel mappings with respect to the similarity metric.

closed-form or by solving a multidimensional optimization problem. The advantage of a closed-form solution is that it can be obtained in finite number of trials using certain elementary functions. Furthermore, the solution is exact (*i.e* not approximated). This can also be its disadvantage, if the data on which the solution is based on is not accurate. Therefore we have to make sure that the feature matches, used for determining the parameters, are accurate, otherwise we have to find them using a search strategy, which leads to the best fit approximation. The registration problem typically comes down to determining the matching transformation parameters by traversing the search space and looking for the maximum of the similarity metric. This procedure equals a multidimensional optimization problem, where the number of dimensions corresponds to the DOFs of the underlying transformation. Basically one need to select the method that is able to find the maximum similarity in the search space, which is not well behaved function in most of the cases. The most straightforward and the only approach that guarantees the global optimal solution is exhaustive search over the whole parameter space. Although it is computationally demanding, it is frequently used if only translation parameters are to be estimated. When this is not feasible, techniques such as Linear Least Squares (LLS) [Lawson and Hanson, 1974]), Non-linear Least Squares (NLLS), Gauss-Newton numerical minimization [Björck, 1996], Gradient Descent method (GDm), Levenberg-Marquardt optimization method [Moré, 1978] and Powell's multidimensional direction set optimization method among others [Powell, 1964], have been successfully applied in the image registration. Which approach is used is in many ways determined by selected feature space.

2.1.5 Similarity metrics

The similarity metrics can be split in two main classes: the mono-modal and multi-modal measures as shown in Figure 2.6. The mono-modal similarity measures are only applicable for registering images of the same modality. In contrast, the multi-modal similarity measures can register different modalities. They are, however, in general inferior to the mono-modal measures. Depending on the intensity mapping of the reference and floating image a different similarity measure must be chosen. This step is very sensitive to image degradation and erroneous feature detection. The choice of similarity metric needs to consider that the features corresponding to the same physical structures can be dissimilar due to different imaging conditions or types of imaging sensors. The similarity metric needs to be invariant to such possible degradation. Moreover it needs to be unambiguous enough to distinguish among different features and stable enough so as not to be influenced by slight unexpected feature variations and noise. Suitable similarity metric selection is closely related to the choice of feature space, since it measures the similarity of selected features. The invariant properties of the image, its intrinsic structure, are extracted by both, the feature space and similarity metric.

Similarity measures, which are grouped into *Identity* mappings, rely on identity relationship between the pixels in the reference $I_R(x,y)$ image and pixels intensities in the floating $I_F(x,y)$. The Mean Squared Difference (MSD) is the simplest mono-modal similarity measure assuming an identity relationship between pixels. It is a simple and robust least squares measure. The Mean Absolute Differences (MAD) is also a mono-modal similarity measure that is similar to MSD but less sensitive

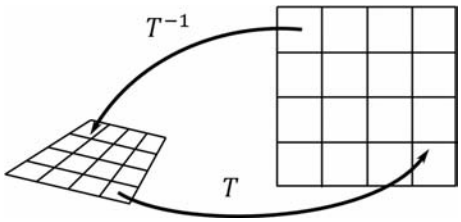


Figure 2.7: Spatial mapping.

to outliers. Metrics such as Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD) also straightforwardly compare raw image intensities. Many medical images are captured with uncalibrated intensities. This is a big problem with all aforementioned metrics. Cross Correlation (CC) and Normalized Cross Correlation (NCC) metrics, in contrast, are capable to compensate a linear relationship. Correlation means measuring the difference between the data and the best fit of a line to the data.

Multi-modal registration, in contrast to mono-modal case, generally has no linear relationship between the intensities in the images I_R and I_F. Thus, the previously mentioned metrics can not be directly applied because the intensities are related by an unknown function $\mathscr{F}$ or statistical relationship which is unknown *a-priori*. It is, however, possible to use the previously presented subtraction and correlation based similarity measures for multi-modal registration by estimating the intensity mapping I_F. This intensity mapping, however, is usually neither smooth nor easily parametrizable. To characterize alignment in such cases evaluating the histogram sharpness is the solution. This forms the basis for Mutual Information (MI) and Normalized Mutual Information (NMI) metrics which incorporate Shannon's entropies of the individual normalized image intensity histograms and are maximized with increasing similarity. A problem using Shannon's entropy for image registration is that a low value can be found for complete misregistration (*e.g.* a case where only one element is in the overlapping area of the two images).

The aforementioned similarity measures differ in the assumed relationship between the intensities of the matched images. Many ways exist for modifying this criteria, *e.g* multi-resolution, multi-scale, edge/boundary/geometric feature extraction. Unfortunately there are no clear rules about how to select a metric, other than trying some of them in different conditions. An application best criteria largely depends on the type of data. Such factors are the difference in the provided information, what contrast is shared and how much they overlap play important roles. In some cases it could be an advantage to use a particular metric to get an initial approximation of the transformation, and then switch to another more sensitive metric to achieve better precision in the final result.

2.1.6 Warping and blending

Once a spatial transformation is known, the input images have to be mapped in to the output image. This process is also referred as *warping* because the resulting image represents a linear or nonlinear *warp*. It is performed either by *forward mapping* or *backward mapping* in conjunction with pixel interpolation as shown in Figure 2.7. During forward mapping each pixel from the input image is transformed $x' = Tx$ and copied to the output image. This, however, does not ensure that the pixel coordinates x' will fall onto exact pixel locations. Moreover, two or more input pixels could be mapped to the very same output pixel and some output pixels might not get a value assigned at all causing mapping gaps. These issues can be avoided by calculating how much area each input pixel occupies in the output image using image sampling. Although the approach produces good results, it is not trivial and computationally expensive. During backward mapping for each output pixel the coordinates in the input image are calculated as $x = T^{-1}x'$ and copied over. As the pixel coordinates in the input image generally do not fall onto exact pixel locations, the pixel intensity/color

is interpolated among the nearest input image pixels. This method is commonly practiced because it is easier to implement and computationally faster than forward mapping. However, it might not be always possible to find the inverse of the transformation to successfully perform backward mapping.

Interpolation techniques used for mapping include Nearest Neighbor (NN) interpolation, which sets the intensity of a given pixel equal to the intensity of the closest pixel. It is very fast and does not introduce new grey values in the result. However, it often heavy artifacts. Bi-linear interpolation determines the grey level from the weighted average of the four closest pixels. It generally produce less artifacts than NN interpolation but it smooths and blurs the image and thus reduces spatial resolution. Another type of interpolation is Cubic Convolution Interpolation. This function is a special type of *approximating function* as it must exactly match with the sampled data at the sample points. Bi-cubic convolution is derived for the 2D case. Cubic convolutions are better at retaining the original intensity values than NN and bi-linear interpolation. However, they are much slower.

Warping result provides an alignment of the images. To achieve a good visualization of this alignment, a suitable *blending* is required. The simplest strategy - α*-blending*, is to add pixel values together using a percentage of the color of each pixel. A smoother result can be achieved by changing the percentage of the blending as the blending is taking place. Such strategy is called *gradient blending*, where a direction of the gradient has to be chosen. Of course applying both the horizontal and vertical blending would result in too much of a blending reduction. The simple solution is to choose the larger of the two reductions to use at any given pixel. In *feathering* approach, the pixel values in the blended regions are weighted average from the two overlapping images. This approach, however, does not tolerate the presence of exposure differences. *Pyramidal blending* downsizes the image into different sizes using the Gaussian function and then expands the Gaussian into the lower level and subtracts from the image in that level to acquire the pyramid representation. This is applied to both images which are then combined in different levels by blending partial images from each of them. There exist a lot more numerous strategies which allow to achieve seamless image fusion such as *multi-band* blending, *dissolving*, *multiply and screen*, *softhard light*, *dodging and burning*, etc. We refer the reader to the theory of photographic tone reproduction for more details [Reinhard et al., 2002].

2.1.7 Classification of registration methods

There exist a number of criteria to categorize and to classify image registration methods. Excellent and elaborated classification can be found in [Maintz and Viergever, 1998, Viergever et al., 2016]. Broadly speaking, there exists two ways to tackle image registration problem. These are either *intensity-based* registration or *feature-based* registration. In addition, several *hybrid* methods, combining the merits of both, have been proposed. Registration approaches can be also grouped with respect to the degree of rigidity the applied transformation intend to recover.

Intensity-based registration Intensity-based approach maximizes a measure derived directly from the pixel intensities. It is based on the assumption that there exists a relationship between the pixel intensities in both images. The registration error in this approach does not depend on feature extraction nor landmark detection, but is at the same time more difficult to discover. A schematic illustration of intensity-based registration is shown in Figure 2.8. These methods work in an iterative manner by warping one image on to the coordinate system of the other using the current transformation estimate, computing an update of the transformation and repeat these steps until a stopping criterion is satisfied (*e.g.* the quality of the fit is less than a predefined threshold). The transformation function in intensity-based registration methods often consists of several terms which model both geometric and photometric transformation. A noise term is also often included in the modeling.

Intensity-based methods are arguably at a disadvantage here. Any differences between the two images that are not accounted for by either geometric or photometric terms in the transformation

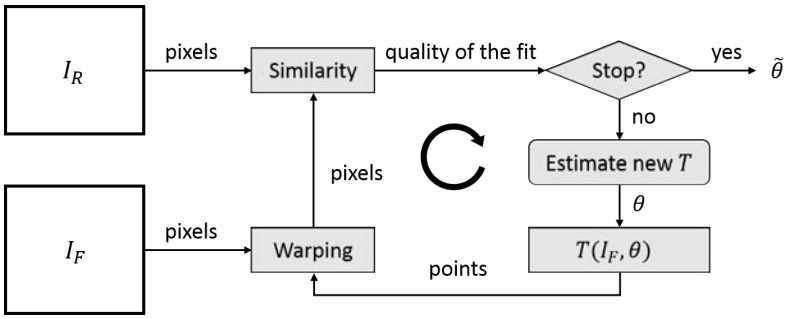


Figure 2.8: Schematic illustration of the intensity-based image registration approach.

function must be absorbed by the noise model and similarity metric. If the photometric model is inadequate, corresponding pixels in the two images may exhibit large differences even when the geometric registration is exact.

Feature-based registration Feature-based approach, in contrast to intensity-based techniques, requires the extraction of image features, where the registration error depends on their localization accuracy. A schematic illustration of feature-based registration is shown in Figure 2.9. Most feature-based methods start by extracting image features and fine tune their mappings to the correlation of these features. The extracted features in one image are matched to the ones in the other image, either by their appearance similarity or by geometric closeness. During the matching, correspondences are formed between features in the two images and a transformation is estimated using an objective function based on a geometric distance.

The starting point is the feature extraction. This combines feature point localization and description. The next step is feature matching to obtain point correspondences which are then used for transformation estimation by applying any of the search strategies discussed earlier §2.1.4. These correspondences, however, most likely contain *outliers*. Outliers degenerate the quality of the transformation parameter estimation. These are either wrongly matched point correspondences or a feature point that is noise or does not belong to the fitted transformation. RANdom SAmple Consensus (RANSAC) [Fischler and Bolles, 1981] is one of the widely used algorithm to reject outliers while jointly estimating a transformation. The idea behind is that if an outlier is chosen to compute the current fit, then the resulting line (*i.e.* the line we are fitting to the data if the LLS method is applied) will not have much support from the rest of the points. The problem with this method is that in many practical situations the percentage of outliers is very high. Overall, a good matching strategy should be able to reduce the number of outliers while preserving the good ones.

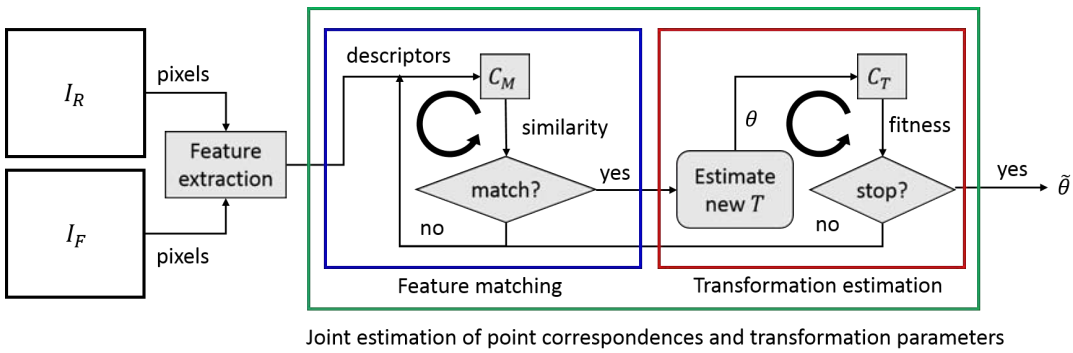


Figure 2.9: Schematic illustration of the feature-based image registration approach.

An important motivation for using features is their invariance to a wide range of photometric and geometric distortions. The localization of intensity discontinuities and auto-correlation maximum, the building blocks of many edge and point feature detectors, is unaffected by large illumination changes. Even though feature-based approaches face challenges in scenes with repeatable textures, it has been observed experimentally that it is feasible to extract meaningful features with high repeatability.

Rigid registration It is often assumed that between image acquisitions, the anatomical and pathological structures of interest do not deform or distort. This *rigid body* assumption simplifies the registration process and rigid registration yields a global rigid transformation of one image to match to another image. Techniques that make this assumption have quite limited applicability. Many organs do deform substantially, for example with the cardiac or respiratory cycles or as a result of change in position. The brain within the skull is reasonably non-deformable provided the skull remains closed between imaging, and that there is no substantial change in anatomy and pathology, such as growth in a lesion, between scans. Imaging equipment is imperfect, so regardless of the organ being imaged, the rigid body assumption can be violated as a result of scanner-induced geometrical distortions that differ between images. Similar situation can be observed in retinal imaging where the eye assumed to remain unchanged within the head of the patient. However, it is not true as the use of contact lens necessary to examine the retina and obtain the SLI modality induces the pressure on the eye ball which may affect the perception of the retina itself.

Non-rigid registration Non-rigid or deformable registration methods take into account more complex deformation by adding more DoFs as it has been discussed in §2.1.4. These methods allow local warping of image features, thus providing support for local deformations. In medical imaging, deformable registration is particularly common in longitudinal studies such as in child development, ageist studies and also in comparisons between controls and pathologies to assess progress or remission of disease. The most popular approaches come in two varieties, some assume brightness constancy in their cost function being optimized while others use information theory based cost functions that do not require the aforementioned restrictive assumption. The former are applicable only to the same modality data sets while the latter can be applied to multi-modal data sets.

2.1.8 Factors complicating image registration

There exist numerous factors complicating the task of image registration both in medical imaging and computer vision in general. They mainly arise from the specifics of the image acquisition process or data driven. The following paragraphs explain several important aspects that make image registration a challenging task.

Data quality and geometric data complexity Low data quality means that points and even whole structures (*e.g.* individual blood vessels) can appear in one image, but be missing in the other. These outliers can cause mismatches that skew the parameter estimates, converting small misalignments into much larger ones. Repetitive structures such as meshes, multiple elongated structures such as blood vessels and nerve fibers, and high-frequency structure such as in brain images create many opportunities for mismatches when there are even small misalignments. Such complexity raises the level of accuracy required in the initial estimate.

Finding suitable features This is currently the main challenge for feature-based methods. Particularly in multi-modal registration, finding the same feature in different types of images is crucial factor. One may think that intensity-based methods, which do not require feature detection, should be easier

to apply. This, however, is not true due to the photometric distortions that affect appearance of the image pixels. Thus, failure in establishing pixel-to-pixel correspondence for intensity-based registration is equivalent to the failure in finding corresponding keypoints for feature-based registration.

Modeling geometric distortion There are many different causes of geometrical distortion that may occur between images to be registered. Underlying physics of the acquisition process vary between different image modalities. Particularly, lens distortion is one of the major factors that induces deviation from the rectilinear projection (*i.e.* a projection in which straight lines in a scene remain straight in an image). It depends on the type of objective lens used in the image acquisition process. There are two major types of lens distortion: radial and tangential. In practice, radial lens distortion is often dominant and generally increases as the focal length decreases.

Another important factor arises from the fact that different types of organs deform in different ways due to the elastic properties of the tissues, blood flow and patient movements. During the retinal imaging process, the object assumed to be rigid and small random amounts of rotation, scale change, and shearing often occur caused by patient movement and eye saccadic motion between consecutive acquisitions. Planar-to-spherical mapping due to the curved retina induces additional distortions. The geometric distortion induced into OCT modality are caused by by refraction, curvature of the intermediate layers up to the depth of interest and the scanning procedure. It is theoretically possible to correct or account for geometrical distortions by accurately modeling them. In practice, however, it may be difficult to get the information needed to do this consistently. Different types of transformation models discussed in §2.1.4 meant to accurately model the assumed deformation. They, however, only from an approximation which always contains certain degree of errors.

Modeling photometric distortion Photometric distortion results in the same tissue in different places appearing in the image with varying intensity. Since the intensity distortion is likely to be different between images (even images of the same modality acquired at different times), this effect results in the intensity mapping being non-stationary (*i.e.* changing over the image). In medical imaging, the effect is common in MRI, where the shading is caused primarily by inhomogeneity, and is also present in radiographs. In retinal imaging the optical vignetting effect in which the image illumination declines as getting away from the camera axis often occurs. Also, light rays reflected from the retina travel through the cornea and a series of optical lenses. This effect, however, can be minimized if the fixed camera systems, such as fundus camera, trough calibration to compensate for path deformation at the cornea. In SLIM, however, the acquisition system contains multiple moving parts which results in various specular highlights. Generally, photometric distortion alter the shape of the objective function and affect the optimization accordingly. This, however, can sometimes be minimized if appropriate assumptions are made about the image and the acquisition process.

Accumulated registration errors In multi-view registration and image mosaicing it is often required to compute the transformation relating any given pair of images. One way to achieve this would be to attempt registration between every possible pair of images in the sequence. But in practice, especially for long image sequences, this is impractical. A more common method is to compute transformations only between temporally consecutive images in a sequence, and then use the rule of composition to obtain the transformation between non-temporally consecutive views. However, this method is prone to *drift* registration error which accumulates when composing sequential transformations over long sequences. Registration or mosaicing drift results in misalignment of important structures between images, such as blood vessels in retinal imaging. Moreover, uncorrected mosaicing drift may cause ghosts to appear creating false structures. This is critical in medical imaging as it may lead to incorrect diagnosis. The ways to minimize the drift are linked to the modeling of geometric and photometric distortions which is itself not an easy task.

2.1.9 Assessment of registration accuracy

Applying a known transformation A simple and generally used approach is to apply a known transformation to an image and then use the registration algorithm to re-align both images. Then, the applied transformation is used as GT. This is commonly suitable for mono-modal registration. Another closely related approach is based on synthesizing images by simulating the imaging acquisition physics and/or material properties and then evaluating the registration algorithm on the synthetic images produced. This, however, may be very difficult due to the complexity of the modeling.

Using similarity metric Since the image registration problem is commonly defined as an optimization problem, an image similarity measure can be used as a crude accuracy measure. Thus, appearance difference between reference and target images can be measured using SAD, SSD, NCC and other metrics alike. These are suitable to evaluate the mono-modal registration algorithms only. NMI and its variations have been proposed specifically to account for the multi-modal image registration. However, most similarity measures frequently used have no geometric/physical significance.

Selecting GT control points A more reliable solution is manually identifying a set of corresponding points in both input images and use them to assess the registration accuracy. These points are called GT control points. To this end the registration error is given in terms of distance between control points and the ones predicted by an estimated transformation model. It has an immediate physical meaning. Root Mean Squared Error (RMSE) is a frequently used measure. A commonly adopted evaluation for registration error in retinal imaging is Centerline Error Measure (CEM) that measures median distances over all corresponding points between two vessel centerline models.

Using object phantoms In some studies phantoms are used to assess the accuracy since they allow accurate control/simulation of the patients' movements. This is common in minimally invasive procedures applied to abdominal organs, for example. In retinal imaging, this is not useful. The Dice Similarity Coefficient (DSC) and Tanimoto Coefficient (TC) quantify the amount of overlapping regions and have also been used to assess the registration accuracy specifically in the context of non-rigid registration.

Visually assessing by an expert Another widely used approach to ensuring acceptable accuracy is visual assessment of the registered images before they are used for the desired clinical or research application. This, however, is a subjective evaluation because different experts may grade the registration result differently.

2.2 Light-related imaging artifacts

An imaging artifact is a detail appearing in an image that is not present in the original object. Light-related imaging artifacts are caused by the reflections of light from surfaces in real scenes or from the parts of the imaging system such as objective and contact lenses. They can be generally classified to diffuse and specular [Nayar et al., 1993]. The diffuse component results from light rays penetrating the surface, undergoing multiple reflections, and re-emerging. In contrast, the specular component originates from the light rays, that are incident on the surface, reflected such that the angle of reflection equals the angle of incidence. This causes artifacts of different degrees which can be broadly grouped as follows.

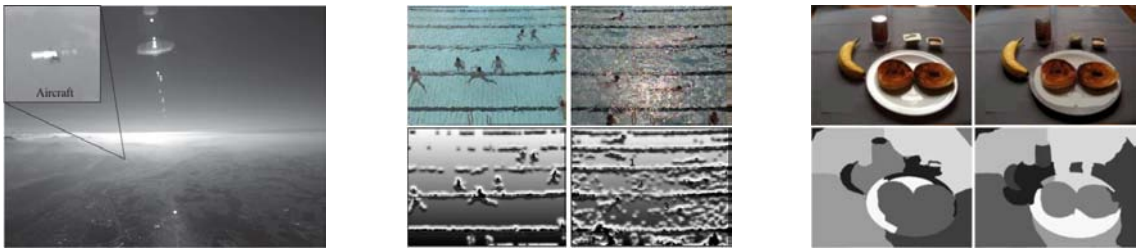
Glares appear as strong bright spots resulting from a light energy concentrated in a compact fold. This causes specular reflections that entirely obscure the image content. They are characterized

as a difficulty of seeing in the presence of bright light such as direct or reflected sunlight or artificial light such as car headlamps at night.

Lens flares are the artifacts occurring in optical lens systems if light is internally reflected or scattered in between the optical elements [Kingslake, 1992]. This usually happens if a bright light source is within or close to the camera field of view. The characteristics of the final artifacts heavily depend on the mechanical and optical properties of the lens system.

Specular highlights are the bright spots of light that appears on shiny objects when illuminated. They sometimes appear to be a spot with diffused edges that gradually get less bright with distance from the brightest spot. Examples of these include the sunlight glinting off ripples in the ocean or an artificial light reflecting from the curved surface of the fruits.

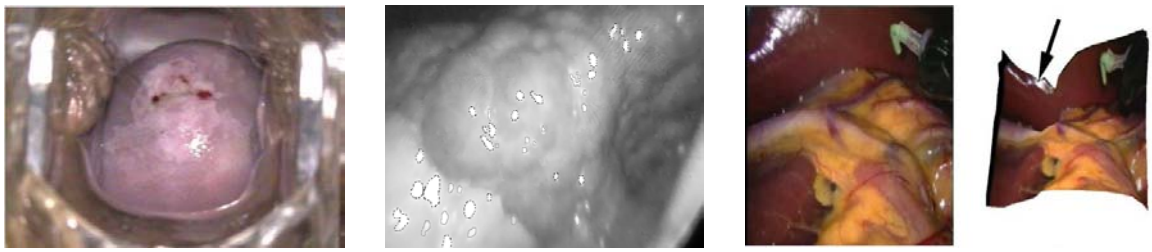
These light-related artifacts play a major role in many computer graphics and vision problems. They provide a true sense of realism in the environment, give a strong visual cue for the shape of an object and its location with respect to light sources in the scene [Collins and Bartoli, 2012] and provide useful cues for object recognition [Osadchy et al., 2003]. However, in most cases, specular highlights are undesirable in images. They are often considered as annoyance in traditional photography. Uncorrected specular reflections cause many coputer vision algorithms such as segmentation and object tracking methods and techniques for shading analysis to produce erroneous results. An image with lens flares taken from an airborne platform shown in Figure 2.10a illustrates how the lens flare can complicate the aircraft tracking [Nussberger et al., 2015]. In the upper left corner a magnified cutout of the traffic aircraft is shown - it looks similar to some of the lens flares. Specular reflections appearing on the water significantly complicate the task of human detection in surveillance system for an aquatic environment [Wang et al., 2004]. Typical scenarios during the day and at night time in the swimming pool are shown in Figure 2.10b with corresponding 3D transformations used for detection in the second row. Various phenomena result from the planar variations of the water surface. Specifically, region 3 shows how specular reflections partly hide a swimmer below the water surface. Specular highlights also affect the performance of image-based dietary assessment with mobile devices [He et al., 2012], as shown in Figure 2.10c. Food image segmentation results shown in the second row are improved after the removal of specular highlights removal.



(a) (b) (c)

Figure 2.10: Some examples of light-related artifacts in non-medical applications.

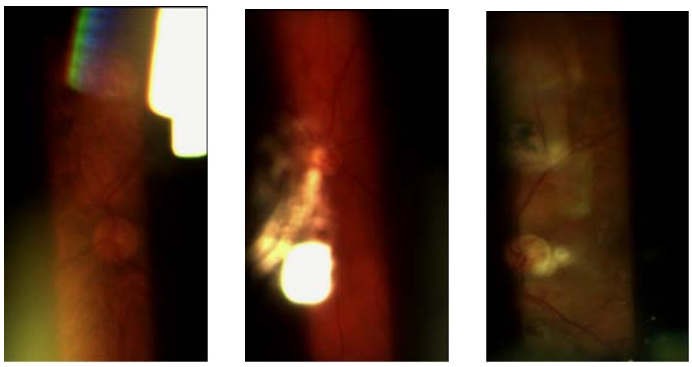
The situation becomes more critical in medical imaging applications. Specular highlights may be very noticeable or just a few pixels out of balance but can give confusing appearances of the photographed object, especially with pathology, that may lead to a wrong diagnosis. Glare eliminates all information in affected pixels and can introduce artifacts in feature extraction algorithms used for computer-aided diagnosis in colposcopy [Lange, 2005], as shown in Figure 2.11a. A cirrhotic liver surface with specular reflections, as shown in Figure 2.11b, complicates classification task [Chakraborty et al., 2014]. Specular reflection in laparoscopic images, as shown in Figure 2.11c, introduce visible errors in the recovered depth information for 3D reconstraction [Stoyanov et al., 2005].



(a) (b) (c)

Figure 2.11: Some examples of light-related artifacts in medical imaging applications.

Often in retinal imaging with a slit-lamp, the imaging system is equipped with multiple lenses. The intensity of the light is controlled and the imaging sensor is not fixed (*i.e.* hand-hold or placed on a moving base). The lenses can be changed during retinal examination with respect to the specifics of the observed pathology. The light intensity may be intentionally varied either in attempt of finding a tradeoff between patient's comfort and the requirements of the imaging protocol. The variation of the light intensity can be also required to obtain different illumination for subsequent image frames. These factors induce specular highlights that are difficult to separate from the retina, as shown in Figure 2.12. If the sensor direction is varied, highlights shift, diminish rapidly, or suddenly appear in other parts of the retina. This poses a serious problem for the image registration methods that rely on image correspondences.



(a) (b) (c)

Figure 2.12: Some examples of light-related artifacts in SLIM.

2.3 Summary

The background information on image registration and its related issues such as mosaicing drift and light-related imaging artifacts are presented in this chapter. It can be derived that image registration is not an easy problem, especially in medical imaging applications. Often, designing an accurate and robust method highly depends on the specific task in hand. Such factors as assumed type of geometric transformation between images, imaging artifacts, and required registration accuracy and application-dependent data characteristics play a crucial role. Both feature-based and intensity-based registration approaches have their advantages and drawbacks. However, feature-based methods appear to be preferable in SLIM.

Chapter 3

Previous Work

In this chapter we give a detailed discussion of the previous work. We determine the precise scope of the literature review first in §3.1. We organize it into separate sections with respect to the thesis objectives. In §3.2 we describe the state-of-the-art in retinal image mosaicing where we separately discuss the case of two groups of retinal image modalities CF, IR, FA together and SLI. In this section we also review the work on the assessment of transformation models applied in retinal image registration and study the methods for reduction/correction of accumulated registration errors which usually occur in the case of mosaicing long image sequences. In §3.3 we review related work on multi-modal medical image registration with an emphasis on methods designed for retinal image modalities. This is followed by an analysis of the methods on detection and correction of unwanted specular highlights in medical imaging as well as in non-medical applications in §3.4.

We accentuate the following important points in the discussion. A lack of comparative evaluation of geometric transformation models within the scope of sequential mosaicing of long image sequences can be observed. This is particularly important if one has to make a choice for the suitable model in such applications. While existing solutions to the problem of drift in generic image mosaicing applications rely on the concept of Bundle Adjustment (BA) and provide satisfactory results, the baseline mosaicing method ignores it. The state-of-the-art methods in specular highlight correction show good performance in generic applications but they are limited to correct only strong glares while leaving aside other degrees of illumination artifacts. Despite the many existing solutions in multi-modal retinal image registration, none of them deal with such a challenging modality as SLIM.

Contents

3.1 Scope **30**
3.2 Image mosaicing **30**
3.2.1 Application to retinal imaging 30
3.2.2 Assessment of transformation models 36
3.2.3 Reduction of accumulated registration errors 37
3.3 Multi-modal image registration **39**
3.3.1 Medical imaging applications 39
3.3.2 Retinal image modalities 41
3.4 Detection and correction of light-related imaging artifacts **43**
3.4.1 Specular highlight correction in medical imaging 43
3.4.2 Specular highlight correction in the non-medical domain 44
3.4.3 Application to retinal imaging 45
3.5 Summary **45**

3.1 Scope

Before diving into the review of the related works and the discussion of their advantages, disadvantages and applicability to our project, we would like to define precisely the scope of the literature we will cover. Because image mosaicing has a very broad application in different domains, exceeding medical imaging and its general framework does not change depending on the application, we focus on the image mosaicing methods applied in retinal imaging only. We, however, provide a comprehensive overview of all the available material that has been published in major journals and conference proceedings in the medical imaging domain. We separately review methods applied to fundus photography and SLIM as there exists a significant gap between the two. Similarly, we limit the review of methods for assessment of the transformation models to ones which were used in retinal image registration and mosaicing only. Regarding the works devoted to the reduction of registration errors, we briefly talk about different approaches including graph-based solutions and variational methods. However, our main focus is on the most representative works based on BA as this registration refinement technique has become the state-of-the-art in drift-related problems. We review both the feature-base and the intensity-based BA methods and specifically highlight the key-frame variations of this technique as we adopt it in our research. Because multi-modal registration found its major application in medical imaging we do not review related methods which can be found in remote-sensing, for example. Instead, we give an overview of the most representative works in multi-modal registration of medical images which spans mainly the last two decades. We categorize them with respect to the complexity of the assumed deformation one aims to recover. We also emphasize a number of interesting works related specifically to registering retinal images and SLIM. Finally, works on detection and correction of light related imaging artifacts are studied in both the medical and non-medical domains, where single-image based solutions are discussed along with multi-view based approaches.

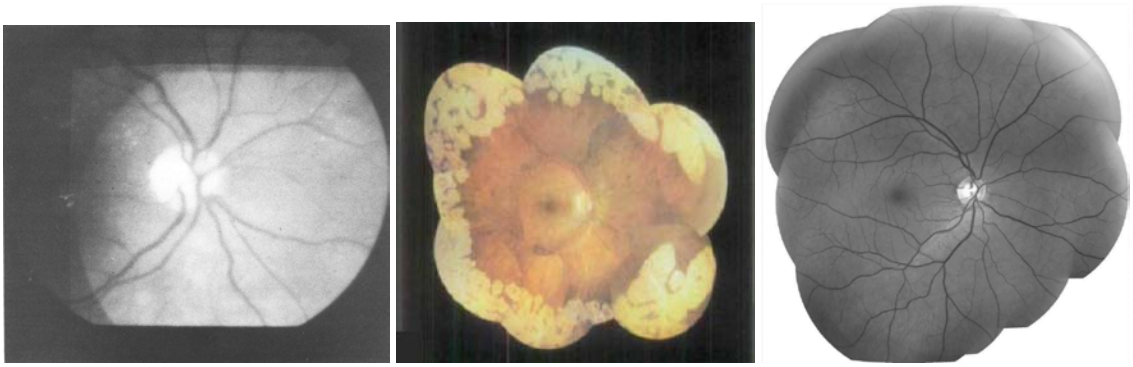
3.2 Image mosaicing

3.2.1 Application to retinal imaging

Excellent various reviews of image mosaicing methods covering different applications can be found in [Irani et al., 1995, Kumar et al., 1995, Abraham and Simon, 2013, Ghosh and Kaabouch, 2016]. In this chapter we mainly focus on the techniques proposed in the retinal imaging domain. Due to various factors, the data obtained from many medical image modalities suffer from a small field of view. By applying image mosaicing in such cases, experts have access to information at a macroscopic scale while retaining the microscopic level of details. This is particularly important for the diagnosis of diabetes related retinal diseases and their treatment planning. Therefore, a lot of research is being carried out for the mosaicing of retinal images.

Mosaicing Fundus Photographs

The majority of existing works on retinal image registration and mosaicing uses images obtained from a fundus camera, either CF, IR or FA. The first attempts for automatic mosaicing of fundus photographs date back to the 80s [Tanaka et al., 1978]. The method adopted initially involved thresholding and then skeletonizing the resulting binary image. Various features including the length of the blood vessels, their direction and the number of branches at branch points were obtained from the skeletonized image. The above procedure was carried out for all images to be assembled. The overlapping images and their relative translation values were found by analyzing the measured features. The analysis of features was restricted to branch points. It is obvious that simple thresholding employed in this method could not ensure the correct extraction of the blood vessel network. Since



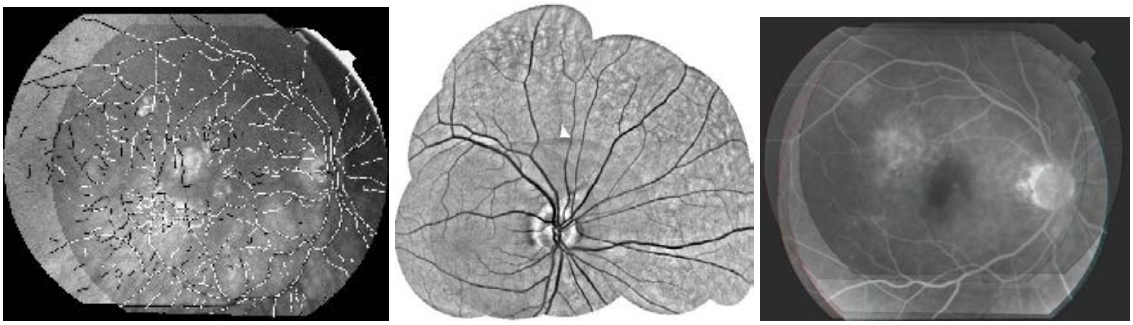
(a) (b) (c)

Figure 3.1: Retinal image mosaicing results reported by (a) [Pham and Abdollahi, 1991], (b) [Mahurkar et al., 1996] and (c) [Can et al., 2002] respectively.

the method was primarily based on the structural details of the vessels, any error in the detection of the contrast variation caused inaccurate assembly. More meaningful results were later obtained by [Pham and Abdollahi, 1991]. The mosaicing method was based on matching only those blood vessels which are common to the overlapping images rather than applying a global registration technique. Also, unlike [Tanaka et al., 1978], the proposed method used all the blood vessel segments in an image rather than just those belonging to the branch points and an angular displacement (rotation) was included in the estimated transformation parameters. In addition, a first attempt to hierarchical refinement was introduced with local matching of segmented vessels followed by adjustments on a global scale. Figure 3.1a shows one of the reported registration results.

In [Mahurkar et al., 1996] individual 45° fundus photographic slides were digitized for creating a wide mosaic semi-automatically. A human operator was required to identify 12 background points by placing the cursor over the region and marking the control point. A small window centered on the control point was used to compute the mean color levels at that point. Data from these control points were used to identify the parameters of the five polynomial models. The best model was used for background subtraction. The blood vessel crossings were also identified manually by the operator as point correspondences between two adjacent images. Data from the nine control points were used to identify the parameters of a two-dimensional polynomial warp. This was the first method where retinal imaging was treated as a case of azimuthal projection of a spherical object and a background subtraction technique was applied to account for intensity variations. The mosaics constructed with this method were made of only 8 images as shown in one of the reported results in Figure 3.1b. Later, it served as a basis for a fully automatic solution proposed by [Can et al., 2002]. The major innovation presented was a linear, feature-based non-iterative method for jointly estimating consistent transformations of all images onto one *anchor* image - mosaic. The authors used bifurcations of the vascular tree on the retinal surface to establish image correspondences from pairwise registration both directly with the *anchor* image and indirectly between pairs of non-*anchor* images. An incremental, graph-based technique was used to construct the set of registered image pairs used in the joint solution. A hierarchical estimation of transformation parameters was introduced here starting with translation using similarity weighted histograms. In the next level an affine transformation was estimated using least-median of squares which was then used to initialize an M-estimator to obtain a quadratic transformation. The mosaics construct with this method were made of up to 20 images. An example of the reported results is shown in Figure 3.1c.

In [Stewart et al., 2003] the authors proposed a different hierarchical approach that substantially reduces the requirements on the initial matching conditions. It uses one or more initial correspondences defining the mapping only in a small area around bootstrap regions. In each of these regions

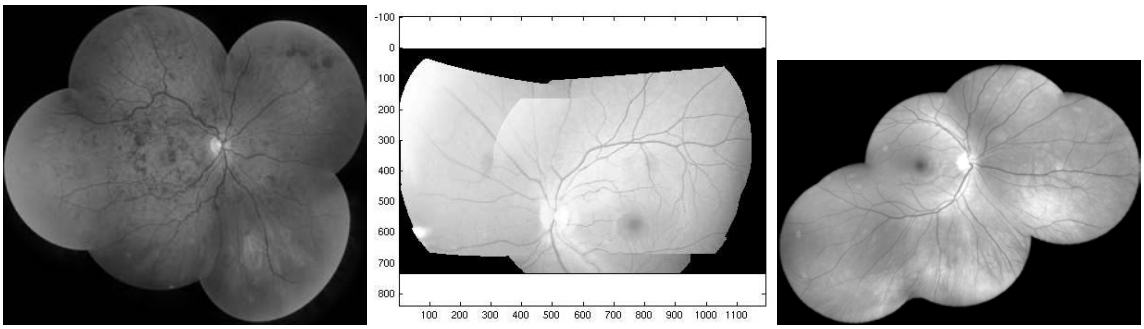


(a) (b) (c)

Figure 3.2: Retinal image mosaicing results reported by (a) [Stewart et al., 2003], (b) [Yang and Stewart, 2004] and (c) [Choe et al., 2006] respectively.

the transformations are iteratively refined using only information from the same area. The region is then expanded and tested to see if a higher order transformation model can be used. The expansion stops when the entire overlapping region of the images is covered. Despite the method was only tested on a temporal pairwise registration of retinal images, its potential to the mosaicing task is obvious. Example of the reported registration result with this method is shown in Figure 3.2a with the indicated vessel features. [Can et al., 2002] produces a final set of transformations for all images if the topology graph is connected. However, this does not ensure accurate alignment, especially when the graph is sparsely connected and there is relatively little overlap between some images. This problem was later addressed in [Yang and Stewart, 2004], where the authors proposed a method to generate new correspondences between such image pairs using the joint alignment transformation estimates and covariance matrices to estimate a more consistent set of transformations. For each pair, transformation parameter covariance matrices are computed and used to estimate the mapping error covariance matrices for individual features from one image. These features are matched in the second image by minimizing the resulting Mahalanobis distance. The generated correspondences are validated using robust estimation techniques and used to refine the estimates. The steps of covariance computation, matching, and transform estimation are repeated for all relevant image pairs until the final alignment converges. The mosaics constructed with this method were made of up to 9 images as shown in Figure 3.2b. The problem of consistency of the resulting mosaic was addressed later in [Choe et al., 2006]. The reference frame that gives the minimum registration error was found by the Floyd-Warshall's all-pairs graph shortest path algorithm, and all other images were registered to this reference frame using an affine transformation model. In this method blood vessel Y-features were extracted using an articulated model and matched across images using the RANSAC method. The mosaics constructed with this method were made of up to 17 FA images. An example of the reported result on registering a combination 6 images FA images is shown in Figure 3.2c. The presentation of other graph-based approach for retinal image mosaicing can also be found in [Aguilar et al., 2007] where the quadratic transformation model was used for registration as well and a new Graph Transformation Matching (GTM) for vessel branch and crossover points was proposed. Example of the mosaic obtained with this method is shown in Figure 3.3b.

All the previously mentioned methods were limited to cases with clearly visible vascular structures. A first step aside from using blood vessel branch points to establish correspondences was made by [Cattin et al., 2006]. The authors proposed a retinal image registration method using SURF to account for images where the vascular tree is not clearly visible. A graph theoretical algorithm was used here to find the *anchor* image that is connected to all the other images through the shortest path and a quadratic transformation model was used for registration as in [Can et al., 2002]. Pairwise registration with this method was evaluated on 100 image pairs and mosaics constructed with this



(a) (b) (c)

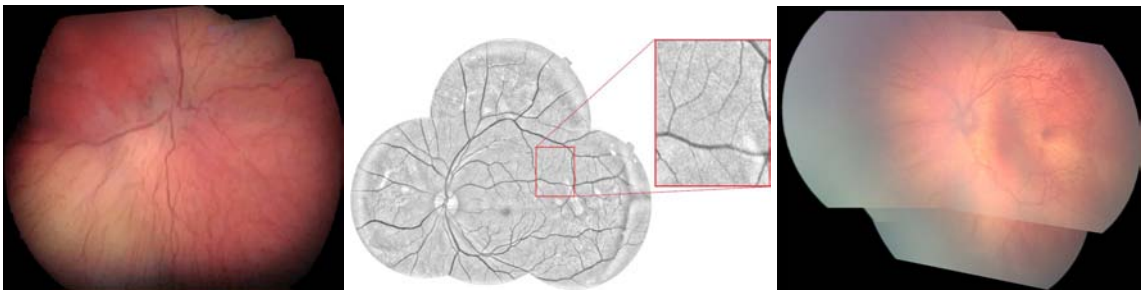
Figure 3.3: Retinal image mosaicing results reported by (a) [Cattin et al., 2006], (b) [Aguilar et al., 2007] and (c) [Lee et al., 2008] respectively.

method were made of only 5 images as shown in Figure 3.3a. Another work which picked up the idea of using local features is [Li et al., 2008a]. The authors proposed to overcome the drawback of SIFT as reported by [Cattin et al., 2006] and presented a method using the *m*-space SIFT. A second-NN matching strategy was used to match the features and quadratic transformation estimation coupled with inlier identification and weighted average blending was applied. The main novelty there was the embedded color information in the SIFT descriptors which gave robustness with respect to color variations.

The majority of previous methods employed the quadratic model derived by [Can et al., 2002] to estimate registration parameters for mosaicing. [Lee et al., 2008] proposed to generate retinal mosaics by a cascading pairwise registration scheme starting from the *anchor* image downward through the connectivity tree hierarchy and a Radial Distortion Correction (RADIC) model was proposed to estimate registration parameters. The RADIC model corrects the radial distortion that is due to the spherical-to-planar projection during retinal imaging. Therefore, after radial distortion correction, individual images can be properly mapped onto a montage space by a linear geometric transformation, *e.g.* affine. The method relies on features obtained from segmented blood vessels and the parameters of the RADIC model are estimated by minimizing the CEM. The mosaics constructed with this method were made of up to 7 images. One of the mosaics reported by this method is shown in Figure 3.3c.

Later work by [Wang et al., 2010] presented a mosaicing methods based on SIFT feature matching and hierarchical transformation estimation with NLLS. The authors employed three transformation models (affine, projective and quadratic) and feathering blending technique is used. The work by [Estrada et al., 2011] uses Gabor filters to obtain retinal features and defines the candidates for matching as the local windows around the maximum filter responses. Naturally, the features detected with this method are concentrated on and around the most prominent retinal vessels originated from the optic disc. The matching strategy here is the minimization of the SSD in a frequency domain and an affine transformation is estimated by L_1-norm minimization. Mosaicing result on CF images obtained with this method is shown in Figure 3.4a.

One of the first attempts to combine intensity-based and feature based registration was presented by [Adal et al., 2014]. The proposed method exploits the intensity as well as the structural information of the retinal vessels. The authors introduced a novel technique to normalize the green fundus image channel for illumination and contrast variation, thereby improving the visibility of the retinal vessels, hence the registration accuracy locally. The method then aligns retinal vessels based on the normalized images. A multiresolution matching strategy was also introduced, where a four level coarse-to-fine Gaussian pyramid is constructed. That is coupled with a hierarchical estimation of the quadratic model [Can et al., 2002] for robust optimization. Mosaicing result on CF images obtained



(a) (b) (c)

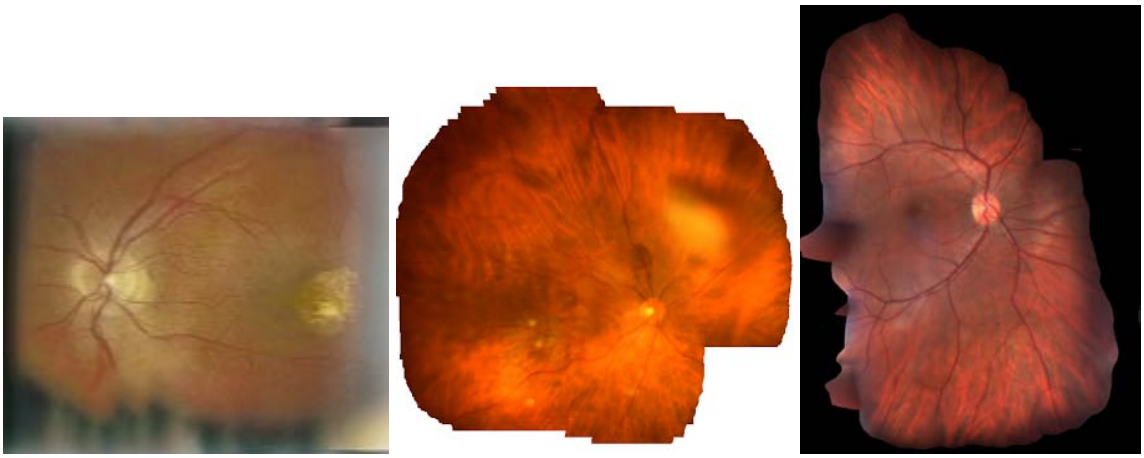
Figure 3.4: Retinal image mosaicing results reported by (a) [Estrada et al., 2011], (b) [Adal et al., 2014] and (c) [Zheng et al., 2014] respectively.

with this method is shown in Figure 3.4b with a zoomed-in region.

The authors in [Zheng et al., 2014] proposed a novel landmark matching formulation by enforcing sparsity in the correspondence matrix and find its solutions which are obtained using linear programming. Their approach relies on two strategies: softassignment and penalization on the to-centroid deviation. The former allows each candidate point to match all reference points while the latter discourages the matched reference points to be scattered. Here the vascular-landmark detection technique [Can et al., 2002] is employed and reinforced Self Similarity (SS) descriptors invariant to global rotation and local affine deformations are introduced. Authors have shown that SS-descriptors have better differentiating abilities compare to descriptors used in [Can et al., 2002] and SIFT-based solutions [Li et al., 2008a, Wei et al., 2009, Wang et al., 2010]. Their performance, however, was not compared to the SURF based methods as [Cattin et al., 2006]. Despite the method was proven to work on the affine, quadratic and the TPS based transformations to register pairs of images, it can fail when the predefined transformation is not well chosen. Example of the mosaic obtained with this method is shown in Figure 3.4c

In contrast to feature-based methods, an intensity-based mosaicing with a mobile low-cost camera was recently presented in [Köhler et al., 2016]. Unlike the previously discussed approaches, the authors exploit video sequences rather than longitudinally acquired images. The main novelty introduced in this method is the super-resolution to obtain multiple super-resolved views to account for subpixel motion that is related to small natural eye movement. These are composed to a common mosaic using intensity-based registration with a quadratic transformation model. Unlike [Can et al., 2002], the parameters of the quadratic transformation were estimated from all the pixels in the image using an Enhanced Correlation Coefficient (ECC) optimization algorithm.

The quadratic transformation model [Can et al., 2002] for retinal mosaicing has a long-term use by many researchers in the field. This, however, was only tested on fundus images. Thus, it raises the question of whether it would be suitable for SLIM too? We address this question further in Chapter 4. Another observation can be made is that one thing shared by all the aforementioned methods is the short length of the image sequence which hardly spans more than 20 images for mosaic composition. The fact that many authors overlooked the case of long image sequences with loops put in question the applicability their approaches to SLIM. Most techniques focus on detecting and matching vascular features *e.g.* branch points, Y-features or crossings among images. These types of techniques are not attainable to real-time mapping using the slit-lamp device because they require a large field of view and high quality images for matching. In addition image registration is not required on-the-fly, hence, the aforementioned methods are not computationally efficient.



(a) (b) (c)

Figure 3.5: Retinal image mosaicing results reported by (a) [Asmuth et al., 2001], (b) [Richa et al., 2014] and (c) [Zanet et al., 2016] respectively.

Slit-Lamp Image Mosaicing

While almost every fundus photograph is valid for the mosaicing, only a small fraction of the frames in raw SLI data is suitable for analysis. Patient preparation and switching from eye to eye result in numerous frames that do not feature the retina. However, the SLI sequence useful for image mosaicing still spans more than 1000 frames for an average imaging session of 2-3 minutes. Mosaicing methods applied to the SLI modality were covered in much fewer works.

[Asmuth et al., 2001] presented the description of the first custom developed software for SLIM. This was an intensity-based approach where pairwise image alignment was accomplished by minimizing the SSD metric between the images with search over translation only. Following alignment and mosaicing of the first two images, subsequent images were aligned with, and added to, the intermediate mosaic in the same manner. In this method, blending is accomplished following decomposition of each SLI into a Laplacian pyramid representation. The final mosaic was then formed by reconstructing the full resolution image from the Laplacian pyramid. Example of the mosaic obtained with this method is shown in Figure 3.5a. Even though the preliminary results reported by the authors were very encouraging, the techniques employed were not suitable to the illumination variations and patient motions verified in practice.

More than a decade after, [Souza et al., 2014] and [Richa et al., 2014] presented a hybrid method for SLIM combining intensity-based and feature based tracking. The method was originally designed for mosaicing intraoperative retinal ophthalmic microscope images [Richa et al., 2012]. An intensity-based template tracking with the SSD metric and a local illumination compensation model was combined with a SURF feature map and employing the RANSAC algorithm. This allowed them to cope with full occlusions caused by uneven illumination and sudden motion. To meet the frame-rate requirements an efficient pixel selection scheme was adopted based on gradient information and inter-frame motion. Efficient Second-Order Minimization (ESM) strategy was adopted to estimate parameters of a rigid body transformation to perform sequential registration. Finally, the weighted averaging blending method proposed by [Szeliski and Shum, 1997] was employed to render a photo-realistic mosaic of the retina. This method was later implemented on the TrackScan platform. Example of the mosaic constructed with this method is shown in Figure 3.5b. Despite of the computational efficiency of the method, it has a number of drawbacks as was previously discussed in §1.1.4. To recall, the accumulation of registration errors over time induces blurring and ghosting effects in addition

to uncorrected illumination artifacts.

Both methods [Asmuth et al., 2001, Richa et al., 2014] lack a global adjustment which leads to drift and misalignments. In addition, segmentation of the slit is based on hard-set thresholds and morphological operators, which fail in more challenging imaging scenarios. Besides, the significantly low image quality and narrow fields of view compared to images obtained with fundus camera in slit-lamp images make intensity-based methods impractical. Recently, [Zanet et al., 2016] proposed a fully feature-based solution for SLIM where the authors account for the problem of uneven illumination and drift. Similar to [Richa et al., 2014], SURF features were used for finding pair-wise translations between frames with RANSAC based estimation. This method benefits from the foreground-aware blending based on feathering that merges video frames into comprehensive mosaics. Unlike other methods discussed above, it also employs a graph-based simultaneous localization and mapping for global BA to achieve consistency in the mosaic. This method is capable of providing a wider FOV compared to [Richa et al., 2014] mainly due to the much better content segmentation technique while it remains unclear whether it improves the registration accuracy because of the absence of quantitative results on registration. Example of the mosaic constructed with this method is shown in Figure 3.5c.

3.2.2 Assessment of transformation models

Despite the variety of works which report on different transformation models for fundus image registration, only a few address their comparison and evaluation. A comparative study of three transformation models (affine, bi-linear and projective) was presented within the scope of multimodal registration [Matsopoulos et al., 1999]. The presented evaluation was performed on 26 image pairs where the values of objective functions averaged over 10 independent executions were compared using two optimization methods for each of the transformation models. The reported results indicate that the affine and the bi-linear transformations appeared both superior in 23 pairs. Specifically, the affine transformation achieved better results than the other transformations in eight pairs, the bi-linear in nine pairs while in four pairs both these transformations performed equally. In the practical implementation of the automatic scheme, the bi-linear transformation model was finally chosen because of its optimal performance in the pathological cases and its property to compensate for more complex deformations than the affine, by using eight independent parameters. Another pursuit to evaluate three transformation models (similarity, affine, and second-order polynomial) was presented in [Laliberté et al., 2003]. A much larger and diverse set of image pairs of different modalities (CF and FA), different resolutions, and different time-captures was used. The reported results on registration performance has been evaluated on 70 image pairs with an overlap-based criterion which indicated that all three transformations were equally good on average for registering the images, although there was a significant number of cases for which a transformation type was better than another one, thus, leaving uncertainty in the choice of the particular transformation model either for multimodal retinal image registration or for SLIM.

A more recent validation for assessing the quality of retinal image registration algorithms and specifically methods in retinal imaging with a fundus camera was presented in [Lee et al., 2007]. The authors aimed at the assessment of *any* retinal image registration method and reported results on similarity, affine, and RADIC models. The main idea of their evaluation strategy is to trace back the distortion path and access the geometric misalignment from the coordinate system of the final registration result (*e.g.* the mosaic). The input mosaic is cut in a "cookie cutter" fashion to create a set of circular images. Then each cut is mapped onto the sphere using an equidistance-conformal mapping through a series of transformations modeled to incorporate the distortions due to the eye geometry and the image acquisition system. The authors also presented a mathematical model that directly converts the mosaic space into camera space and applied a gradual intensity variation to the distortion stage to simulate optical vignetting effect specific to the CF modality. The proposed validation process restores the montage coordinates and compares the registration results to the known

ground truth to assess accuracy. The authors placed a set of landmarks which are in the crossing point set of virtual grid lines on the original montage space. As the image, the coordinate elements of the evaluation point set were modified by the transform matrix for each specific step and compared to its original coordinate elements in the restored coordinate system by computing the MAD of the point displacement vector. A two-step distortion model was specifically derived for this purpose from a simplified modeling of the interaction between the calibrated fundus camera and the eye as a camera rotating around a stationary eye for a limited angular range. Such modeling is difficult to apply to the case of SLIM due to its specificity and inability to calibrate the optical set-up used to obtain SLI.

It is evident that none of the aforementioned evaluation methods has considered the mosaicing of long image sequences obtained in a closed loop motion which is typical for retinal examination with the slit-lamp. Thus, they do not address the problem of accumulated mosaicing drift.

3.2.3 Reduction of accumulated registration errors

The generic registration methods and multimodal registration techniques are designed to compute the transformation between pairs of overlapping images. In situations featuring many views of a scene it is often required to compute the transformation relating any given pair of images, such as in mosaicing methods discussed above. One way to achieve this would be to attempt registration between every possible pair of images in the sequence. But in practice, particularly for long image sequences, this is unattainable. A more common approach is to compute transformations only between temporally consecutive images in a sequence, and then use the rule of composition to obtain the transformation between non-temporally consecutive views. Many aforementioned mosaicing algorithms including the one implemented on the TrackScan platform [Richa et al., 2014] follow this approach. However, it is prone to *drift* - registration error that accumulates when composing sequential transformations.

Various authors have proposed methods for reducing the effect of accumulated drift. Breaking the sequence into smaller sub-sets of frames which are used to create sub-mosaics was suggested by [Mann and Picard, 1995]. The sub-mosaics are then registered and combined to form the final result. [Davis, 1998] solves a linear system, derived from a redundant set of pairwise projective transformations, so as to minimize an algebraic residual defined over the actual transformation matrix elements. Although simple and easy to implement, the algebraic residual used does not correspond to any meaningful geometric error. [Sawhney et al., 1998] proposed a scheme in which the mosaic image is updated one frame at a time, and each additional frame is registered with and blended into the current mosaic image. Despite these are all very practical methods, they are sub-optimal. The optimal solution, as has been known to photogrammetrists for many years [Slama et al., 1980], is to use BA. It is an iterative optimization method, implementing non-linear least squares, computing the mean of the likelihood or posterior distribution (depending on whether prior knowledge is present or not), and taking advantage of sparsity in the system information matrix to speed up its inversion. [Triggs et al., 1999] provided an excellent survey on theory of BA methods.

[Hartley, 1997] described a feature-based BA scheme for the estimation of projective transformations (homographies). He extended the feature-based method for two-view homography computation to the case of simultaneous estimation of homographies over N-views, referring to [Slama et al., 1980] to explain how block matrix methods may be employed in order to render the required nonlinear optimization tractable. This method is guaranteed to produce globally consistent transformations. It is shown to be efficient and accurate, but the problem of how to match corresponding feature points across many views is not addressed. A transfer of BA from photogrammetry to feature-based image mosaicing with application to long image sequences can be found in [McLauchlan and Jaenicke, 2002]. A problem of mosaicing not video sequences but sets of widely separated, uncalibrated still images was considered in [Brown and Lowe, 2007]. Their method used SIFT features to perform wide-baseline matching, and automatically align them into a panorama, optimizing over the whole set for global consistency using BA. In [Yao, 2008], the author exploited a deformation vector propagation

algorithm in the gradient domain to reduce the intensity discrepancy between the composed images. Similarly, a BA algorithm along with a modified-RANSAC algorithm capable of developing a probabilistic model was used in [Li et al., 2008b] to eliminate registration error and make the matching process more accurate.

A fewer attempts at global consistency using BA have also been made using intensity-based methods. [Sawhney and Kumar, 1999] described an analogous to [Hartley, 1997]'s scheme whereby, having obtained initial pairwise consecutive transformations, additional registration is performed between image pairs which are deemed to be spatially adjacent (*i.e.* they capture the same part of the scene). Transformations between any pairs are obtained by concatenating along the shortest-path through the view-graph so formed. Although this method can improve consistency to some extent, the transformations are still computed using a pair-wise algorithm, hence global consistency is not achieved. [Zelnik-Manor and Irani, 2000] proposed a scheme which does aim to simultaneously optimize all transformations, though their method is limited to small perturbations of the camera, since it uses the local bi-quadratic approximation to the full projective motion model. In contrast to BA-based methods, an intensity-based total variational optical flow approach was recently investigated by [Ali et al., 2016] for the case of mosaicing long image sequences in medical imaging. While having improved accuracy and robustness of the TV-L_1 approaches, the authors did not account for minimization speed of the variational energy and did not compare the performance of the algorithm to BA-based approaches, hence, making it inconclusive to the general case of long image sequences.

In real-time systems, BA has been left as a post-processing step for a long time. However, in the past few years a number of real-time local BA-type refinement methods were proposed, which allow one to achieve a similar accuracy to conventional BA while reducing computational cost. [Steedly et al., 2005] explicitly consider questions of computational cost in efficiently building mosaics from long video sequences, though not arriving at real-time performance. The key to efficient processing in their system is the use of automatically assigned key-frames throughout the sequence - a set of images which roughly span the whole mosaic. Each frame in the sequence is matched against the nearest key-frames as well as against its temporal neighbors. Some relatively recent approaches have attempted to combine this idea with probabilistic and statistical techniques in global image registration. Probabilistic filtering method such as [Konolige and Agrawal, 2008] for global image alignment is similar to SLAM in the field of mobile robots. The system state vector consists of stacked parameters. In [Civera et al., 2009], another relatively fast solution, the system state vector is composed of the last camera pose and all map[ed features. Only the correlations among different map features are maintained. But the correlations among different camera state parameters are not. Therefore, the global consistency is not achieved completely. A consistent error still persists after several loops. In contrast, the system state vector in [Xu, 2013] consists of all global transformations parameters corresponding to all images. Other relatively fast methods can be found in [Klein and Murray, 2007, Mouragnon et al., 2009, Lovegrove and Davison, 2010].

The graph-based approaches as [Can et al., 2002, Marzotto et al., 2004, Choe and Cohen, 2005, Aguilar et al., 2007] appear to be impractical within the scope of SLIM due to the considerable complexity of computations. To ensure the consistency of the resulting mosaic given an arbitrary set of images it is necessary to build a maximally connected graph. Even constructing the graph incrementally has a high complexity of $O(N^2)$. This is not a major concern for mosaicing fundus photographs because typically fewer than 20 images are combined in each mosaic. In mosaicing scenarios, where the image sequence spans more than 1000 frames like SLIM, however, it becomes a main drawback. In [Choe and Cohen, 2005], a global registration is introduced to automatically reduce the drift across color and fluorescein images. Global registration is intended to identify the best registration among every pair of images, while minimizing the global registration error using a Minimum Spanning Tree (MST) approach. However, the problem of selecting the reference frame for the mosaic is not addressed. Besides, it does not guarantee the lowest registration error because MST does not consider

the error between the reference frame and other frames. A graph-based SLAM-like method proposed by [Lu and Milios, 1997] was recently employed by [Zanet et al., 2016] within the scope of SLIM. For each frame pair, the transformation estimation was bundled into a linear system of equations. Providing an alternative to the method's complexity $O(N^3)$ that makes it unsuitable for real-time execution, the global refinement was achieved offline for the position of all frames. [Linhares et al., 2016] presented another approach using intensity-based non-rigid fine adjustment. It is an offline procedure that consists of minimizing the pairwise SSD metric in an evenly spaced overlapping set of images. The authors rely on the TPS model to warp the image in a set to a global frame. To prevent excessive image deformation, which may often occur in TPS-based mapping, the authors vary the values of regularization parameters as a function of the NCC between each pair of images. A main drawback of the method is the cropping operation that is applied to the list of image pairs to avoid low-overlapped cases. This eventually leads to an excessive cropping.

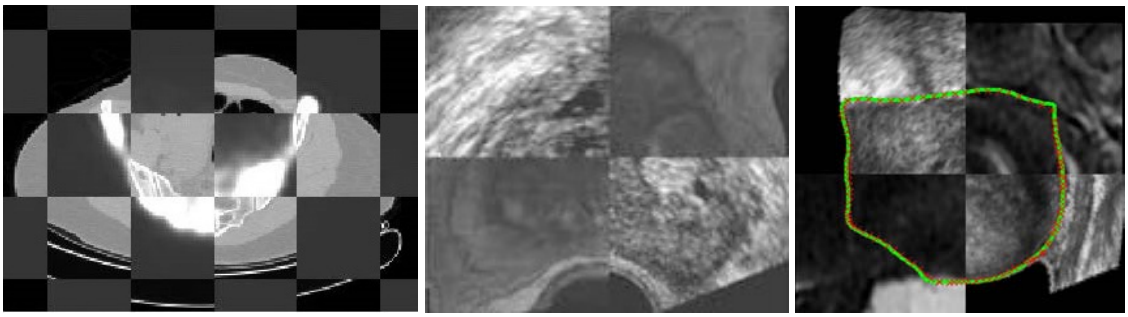
3.3 Multi-modal image registration

Multi-modal image registration is an important aspect of medical image analysis. Different modalities, such as Computed Tomography (CT), PET, Single-photon Emission Computed Tomography (SPECT), FA, CF, Transrectal Ultrasound (TRUS), Transvaginal Ultrasound (TVUS), MRI, fMRI and many others, show unique tissue features at different spatial resolutions. Video images are often acquired during surgery, for example using endoscopes or microscopes. For the purpose of image guidance, it can be useful to relate the video images to preoperatively acquire diagnostic images. Whether registering images across modalities for a single patient or registering across patients for a single modality, registration is an effective way to combine information from different images into a normalized frame of reference.

3.3.1 Medical imaging applications

Numerous attempts have been made to solve multi-modal registration problems in the medical imaging domain. Most of the algorithms are focused on anatomical image modaiities such as MRI, CT, SPECT, FA, functional image modalities fMRI, PET, ultrasound, TVUS, TRUS and optical imaging. These are used in imaging of the brain [Roche et al., 2001], chest [Rueckert et al., 1999, Oktay et al., 2015], abdomen [Müller et al., 2011], pelvis [Mitra et al., 2012, Yavariabdi et al., 2013] and bones [Livyatan et al., 2003, Tang et al., 2006]. The registration approach based on maximizing MI developed by [Maes et al., 1997] became common practice in both rigid and deformable image registration. It is an entropy-based measure of information that an image contains about another image. Later, a normalized version NMI has been proven to achieve better results [Rueckert et al., 1999]. Readers are referred to a survey [Pluim et al., 2003] for more details. Generally, related methods can be categorized with respect to the degree of rigidity assumed in the deformation field.

By exploiting the assumption that a large fraction of the scene is rigid and therefore a single image-to-image transformation function is appropriate, rigid registration approaches aim to vary the registration parameters and search by the optimizer to arrive at a global transformation that gives maximum similarity between registered images. According to the literature, a rigid geometric transformation is mainly applied in two situations. One is the registration of rigid structures on multi-modal images. Thus, a gradient-based 2D-3D rigid registration of fluoroscopic X-ray to CT bone images was proposed by [Livyatan et al., 2003]. [Tang et al., 2006] proposed to register CT and SPECT images of bones using the MI similarity measure. Example of the registration result obtained with this method is shown in Figure 3.6a. Combining intensity and gradient information a registration of 3D ultrasound with brain MRI was addressed in [Roche et al., 2001]. The other scenario is the pre-registration before a more complex geometric transformation [Hellier and Barillot, 2004, Oktay et al., 2015]. After



(a) Bones (b) Prostate (c) Pelvis

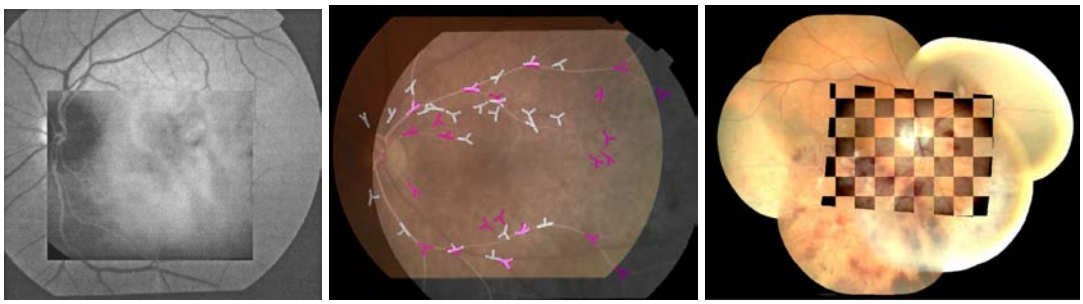
Figure 3.6: Multi-modal medical image registration results reported by (a) [Tang et al., 2006], (b) [Mitra et al., 2012] and (c) [Yavariabdi et al., 2013] respectively.

rigid registration, [Hellier and Barillot, 2004] estimate a deformation field by minimizing a cost function, composed of two terms: the MI measure and a regularization term in order to ensure spatial coherence of the deformation field. [Oktay et al., 2015] presented a block matching approach is used to establish spatial correspondences, where the NCC is used as a measure of similarity. To correct for the residual misalignment, due to cardiac and respiratory motion, between target and source, B-spline based non-rigid alignment follows the global rigid registration.

Most approaches for medical image registration are based on curve or deformable transformations, since the almost all anatomical parts, or organs, of the human body are, in fact, deformable structures. Regarding deformable/non-rigid registration, a comparison of several commonly used algorithms in brain imaging was addressed in [Klein et al., 2009]. Basically, two kinds of curved deformations have been used in medical image registration: free-form transformations, in which any deformation is allowed; and guided deformations, in which the deformation is controlled by a physical model that has taken into account the material properties, such as tissue elasticity or fluid flow. Thus, [Mitra et al., 2012] described non-rigid registration approach for multi-modal images of the prostate. Example of the registration result obtained with this method is shown in Figure 3.6b. In [Yavariabdi et al., 2013] a variational one-step non-rigid Iterative Closest Point (ICP) method was proposed by [Besl and McKay, 1992] for the mapping of small endometrial implants using TVUS/TRUS and MRI modalities. The reported performance shows the methods superiority over TPS-based registration approaches. Example of the registration result obtained with this method is shown in Figure 3.6c. The task of non-rigid registration and fusion of PET and MRI breast images has been addressed by [Rueckert et al., 1999] and [Rohlfing et al., 2003] where a free-form deformations with NMI minimization has been proposed.

The strength of MI so widely applied in related works lies in the fact that it does not presume any functional relationship between the intensities on two images. The intensity relationship is not known until it is estimated during the image registration process. Thus, MI has a broad range of applications and can handle a wide variety of imaging modalities. However, for the very same reason, the optimization is sensitive to the overlap area and is subject to multiple local minima. Therefore, the initialization must be good for it to converge to the correct transformation. Although the NMI partially solves these issues, it is not well applicable to images with long and thin structures such as retinal images, or to the combination of CT and ultrasound images.

The above approaches assume prperties characterizing good image alignment (*i.e.* MI or specifically designed similarity metric), but do not learn them from data. More recently, learning-based approaches to measure image similarity have been proposed. These are particularly deep learning based techniques which learn a similarity measure alone or jointly with the transformation model [Cheng et al., 2016, Simonovsky et al., 2016]. Some methods avoid a complex similarity measure



(a) (b) (c)

Figure 3.7: Multi-modal retinal image registration results reported by (a) [Matsopoulos et al., 1999], (b) [Choe and Cohen, 2005] and (c) [Broehan et al., 2011] respectively.

by applying image synthesis for the source/target image to change the task to mono-modal registration [Van Nguyen et al., 2015]. However, the registration performance then heavily depends on the synthesis accuracy. [Gutiérrez-Becker et al., 2016] proposed to learn a multi-modal similarity measure using a regression forest with Haar-like features combined with a prediction model for a low-dimensional parametric B-spline model. In contrast, [Yang et al., 2017] predicts the initial momentum of the shooting formulation of Large Deformation Diffeomorphic Metric Mapping (LDDMM) [Vialard et al., 2012], a non-parametric registration model and jointly learns a multi-modal similarity measure from image-patches without requiring feature selection.

3.3.2 Retinal image modalities

Multi-modal retinal image registration is an established and ongoing research field. Mostly the intention is to register single retinal image pairs. These are either retinal images of different modalities (*eġ˙*, CF and FA) or images of the same modality taken at different points of time (temporal registration). Multi-modal registration with SLIM is a difficult task due to the significant geometric and photometric differences. The conventional approach to this problem is to detect anatomical landmarks in both images and apply a matching algorithm followed by outlier rejection. Predominantly, algorithms used for retinal image registration are based on a segmentation of the vessel tree and/or extraction of significant vessel features (*eġ˙*, vessel bifurcations). The first automatic registration following this approach was proposed in [Matsopoulos et al., 1999]. The blood vessels were first segmented on both image modalities to obtain binary images. A correlation coefficient adopted for the binary case was used as an objective function in the optimization stage with a Genetic Algorithm (GA) [Goldberg, 1989] to estimate the parameters of a bilinear transformation. Example of the registration result obtained with this method is shown in Figure 3.7a. Later, [Choe and Cohen, 2005] proposed a registration method based on vessel landmarks for CF and FA images of the retina, where a major accent was made on the method to extract Y-features similar to [Can et al., 2002]. The method consists of three main steps: first, seed positions of Y-feature are computed using a Principal component Analysis (PCA)-based analysis of directional filter responses. Second, an articulated model of the Y-feature is fitted to the image features using a gradient descent method. Third, the extracted Y-features are matched by maximizing the MI, and images are registered using an affine model using RANSAC and a global graph-based refinement. While extracting Y-features using an articulated model provides robust, accurate and fully automatic registration, it is only guaranteed to succeed if these features can be reliably identified on both modalities. Example of the registration result obtained with this method is shown in Figure 3.7b.

The ICP is widely used for both mono-modal and multi-modal image registration due to its robustness, simplicity and fast execution time. This iterative method utilizes the nearest neighbor rela-

(a) (b) (c)

Figure 3.8: Multi-modal retinal image registration results reported by (a) [Chen et al., 2010], (b) [Ghassabi et al., 2013] and (c) [Hernandez et al., 2015] respectively.

tionship to assign correspondences between the candidate points of the two images at each step. An extension of ICP to the case of retinal images has been specifically designed Dual-bootrsrap Iterative Closest Point (DBICP) [Stewart et al., 2003], as discussed in the previous section, where region bootstrapping was used to overcome issues such as initialization sensitivity, small overlap, and unreliable matches. A generalized and improved version was introduced later Generalized Dual-bootrsrap Iterative Closest Point (GDBICP) [Yang et al., 2007]. Despite the demonstrated success, GDBICP does have limitations. It cannot handle extreme appearance differences between image pairs and while incorrect alignments in the repetitive region may appear accurate, these produce inconsistent matches between images. These algorithms are often preferred in mutimodal retinal registration. However recent advances propose alternative strategies [Broehan et al., 2011, Chen et al., 2010, Ghassabi et al., 2013, Hernandez et al., 2015].

A method designed for the initialization of a real-time registration procedure for the subsequent video frames of scanning digital ophthalmoscope and the composite image is described in [Broehan et al., 2011]. The authors emphasize that optic disc detection and localization help to accurately estimate the scale parameter for the global alignment with the quadratic transformation model. Example of the registration result obtained with this method is shown in Figure 3.7c. PIIFD was specifically designed for retinal images, where the goal is to register poor quality multi-modal retinal image pairs. The authors use an adaptive transformation to register the image pairs depending on the number of matches available. Example of the registration result obtained with this method is shown in Figure 3.8a. In [Ghassabi et al., 2013] an improvement over PIIFD was proposed where the authors use UR-SIFT features coupled with PIIFD descriptors on retinal images of different modalities to enhance the correspondences. Example of the registration result obtained with this method is shown in Figure 3.8b. The authors of [Hernandez et al., 2015] propose a method to register different types of retinal image modalities using salient line structures extracted with a tensor-voting approach, which are then compared with a Chamfer distance, and the pairwise rigid transformation is estimated. The ICP approach is used to refine the rigid transformations, and a chained-registration is then used to recover in case of wrong pairwise alignment. Final registration is performed using TPS. Example of the registration result obtained with this method is shown in Figure 3.8c.

Because the majority of the proposed solutions does not provide publicly available implementation, it is difficult to assess their performance in the case of SLIM and FA registration. The direct application of ICP and DB-ICP does not produce acceptable results and often fails as we discovered through experiemtns that will be presented in detail in Chapter 7. Moreover, the aforementioned methods have two limitations: they require a detection of feature points in both image modalities and do not account for the accurate registration needed in the macular area, which is difficult to achieve due to the absence of strong features. In [Markaki et al., 2009, Matsopoulos et al., 2004] the

(a) Colposcopy (b) Thoracoscopy (c) Laparoscopy

Figure 3.9: Some results on specular highlight correction in medical imaging reported by (a) [Lange, 2005], (b) [Saint-Pierre et al., 2011] and (c) [Allan et al., 2013] respectively.

authors presented an unsupervised learning algorithm where the properties of Self-organizing Map (SOM) [Kohonen, 1998] were studied and adopted to the problem of both mono-modal and multi-modal retinal image registration. This algorithm has several attractive features useful in our challenging task. It does not require interest point detection and feature extraction in both images. It preserves the topology of the input space. In addition, it is an unsupervised type of learning where the training phase is not time-consuming. It has proved to be strongly resilient to outliers, less sensitive to control parameter selection, and less exposed to the effects of multi-modality and local optima.

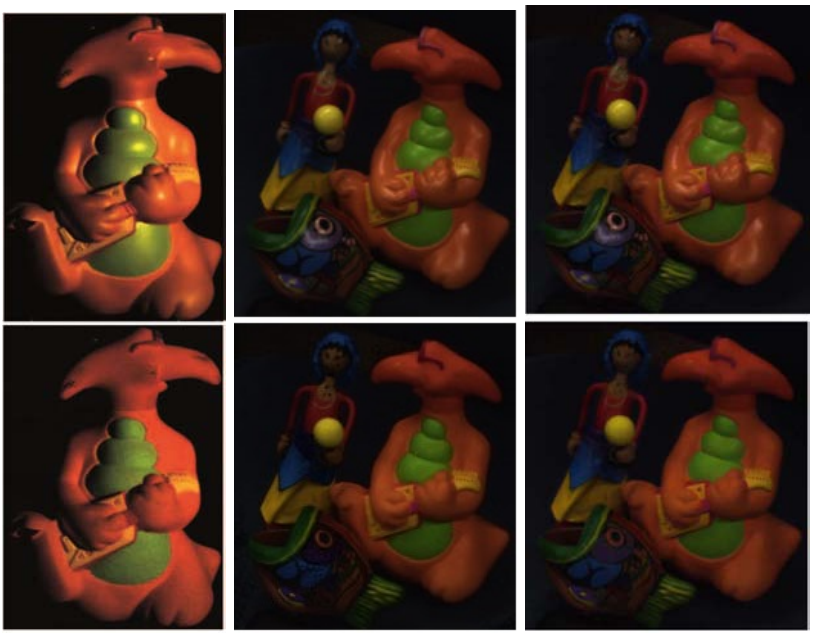
3.4 Detection and correction of light-related imaging artifacts

Numerous methods have been proposed in the literature to solve the problem of a particular type of light related imaging artifacts to some degree. Referring to the types of illumination artifacts that have been discussed in previous Chapter **??**, we divide the detection and correction methods into two categories. First, we consider work that has been presented in medical imaging. This is followed by the review of the methods applied to non-medical domain. The specific case of SLIM is discussed afterwards.

3.4.1 Specular highlight correction in medical imaging

Automatic glare removal in colposcopy was the focus of [Lange, 2005]. The authors proposed a single-image technique where they used the green image component as the feature image, given its high glare to background ratio. Glare regions were detected as saturated regions by adaptive thresholding and morphological top hat filters. The watershed segmentation was then applied to find the contour of the glare regions, which were then restored using inpainting. Example of the glare-free image obtained with this method is shown in Figure 3.9a. A similar approach was presented in [Saint-Pierre et al., 2011] to automatically detect and correct specular reflections in thoracoscopic images. The reported results proved both methods to be adequate on the application specific datasets, one example of which is shown in Figure 3.9b. Specifically, the inpainting was acceptable due to the homogeneous texture of the affected regions while in SLIM this would rather produce false details than restore the 'true' content.

A Machine Learning (ML) method was successfully applied to the segmentation and tracking of surgical tools in laparoscopic videos [Allan et al., 2013]. The authors did not tackle the problem of specular highlights directly. Instead they used a random forest classifier trained with feature vectors which combine color and structural information and relied on their distinctive power. This provided acceptable results as can be seem in Figure 3.9c. In [Chhatkuli et al., 2014] specular highlight segmentation was addressed as a part of ML-based organ segmentation. The specularities were found by



(a) (b) (c)

Figure 3.10: Some results on specular highlight correction in non-medical applications reported by (a) [Tan and Ikeuchi, 2005], (b) [Yang et al., 2010] and (c [Kim et al., 2013] respectively.

thresholding the luminance and saturation channels in the HSV color space after Gaussian smoothing, and the binary decision was appended to the main classification framework.

3.4.2 Specular highlight correction in the non-medical domain

[Tan and Ikeuchi, 2005] and [Shen et al., 2008] are the popular single-image generic solutions nowadays. More recent works on the subject were also presented in [Yang et al., 2010, Yang et al., 2011, Kim et al., 2013, Xu et al., 2015, Nussberger et al., 2015, Sun et al., 2016]. [Tan and Ikeuchi, 2005] proposed a method to separate the diffuse and specular reflection components using chromaticity-based iterations with regard to the logarithmic differentiation of the Specular-free image (SF) image using two spatially adjacent pixels. Example of the reported result obtained with this method is shown in Figure 3.10a. [Shen et al., 2008] separate reflections in a color image based on the error analysis of chromaticity and an appropriate selection of color for each pixel by solving for the dichromatic reflection model as a least-squares problem. Both methods use the concept of SF image, which allows them to decide for specular and diffuse pixel candidates. According to our experiments the direct application of their methods in SLIM is not sufficient. This will be discussed in more detail in Chapter 6.

A generic single-image based solution in [Yang et al., 2010] relies on the observation that the maximum fraction of the diffuse color component in local patches changes smoothly. The authors applied a low-pass filter so that the maximum diffuse chromaticity values can be propagated from the diffuse pixels to the specular pixels. Unlike other methods, this can process high-resolution images at video rate, which makes it suitable for real-time applications. Example of the specular/diffuse component separation obtained with this method is shown in Figure 3.10b. In [Kim et al., 2013], similar to [Tan and Ikeuchi, 2005], the authors derived an SF image by applying a dark channel prior and used it along with a maximum a posteriori probability estimation to separate the specular and diffuse components. Reported results were evaluated only visually and seem to provide a slightly better outcome compared to [Tan and Ikeuchi, 2005] and [Yang et al., 2010] but weaker in computation time.

Example of the reported result obtained with this method is shown in Figure 3.10c.

A solution for specularity removal in stereo-vision was proposed in [Yang et al., 2011], using two images to compute a vote distribution for a number of illumination chromaticity hypotheses via correspondence matching. The authors use motion cues assuming that highlights on the two images do not spatially overlap. Thus, the diffuse component of a pixel in the highlight can be recovered by finding its corresponding pixel in the other view. This assumption holds for the cases where the observation system and light source move independently from each other and the observed object remains static. In SLIM, however, a larger set of observations is necessary to recover the 'true' color while practically it might be possible to obtain a suitable approximation only. An analogous assumption was employed in [Xu et al., 2015, Sun et al., 2016, Nussberger et al., 2015].

3.4.3 Application to retinal imaging

The mosaicing method proposed by [Estrada et al., 2011] is the first to address the problem of imaging artifacts (white spots, speckles, distorted colors) in color fundus images obtained with an indirect ophtalmoscope via directional local contrast filtering and HSV color space based color adjustment in distorted areas. Because this method is designed for specific nature of imaging artifacts originated in the process of indirect ophtalmoscopy, it is not suitable for the case of SLI. [Asmuth et al., 2001, Richa et al., 2014] exploit an intensity thresholding based segmentation of the illuminated retina using different color channels of a single RGB image. This approach is sufficient for simple video sequences where a care over reflection was taken by the ophthalmologist and the patient was not very photosensitive, resulting in reduced apparent motion. However, this is not always the case in practice and a more complex solution is necessary to achieve acceptable results. Recent work [Zanet et al., 2016] employs ML with training on a manually labeled database for per-pixel classification. As in [Allan et al., 2013], the authors opted to use multiple color spaces as features, and added the spatial information. The reported results outperform those from [Richa et al., 2014] and provide robust filtering of strong specular highlights. However, as can be seen from the experimental outcomes, the significant part of the retinal content covered by semi-transparent highlights and lens flares appeared to be excluded from the mosaic, leading to a loss of valuable information.

Most of the aforementioned single-image solutions are capable to correct strong glare. However, they share the same problem: they generally result in noticeable artifacts when applied directly in SLIM. Multi-image methods utilize the motion cues for highlight localization and correction. In SLIM, due to the specifics of the imaging set-up, the apparent motion of specular highlights can be noticed, but, unlike in previous work, more than two consecutive observations are required. Moreover, the limited FOV of one frame cannot capture the highlight fully. Therefore, the motion cues are useful but shall be engaged as soft constraints. Learning appearance variation from multiple images has proved to outperform simpler methods [Zanet et al., 2016]. However, the inability to model complex color and intensity variation of the reflections associated with lens flare make it unsuitable for our goals in SLIM.

3.5 Summary

In this chapter we provided a detailed overview of the methods related to our thesis objectives. It can be observed that developing an image registration algorithm is a complicated task, where all constituting parts have to be designed carefully, considering compatibility among them and suitability to the specifics of the problem being solved. But examples of image registration algorithms in this chapter show that solutions yielding good results exist and are possible to develop through a careful design. Similar conclusion can be made regarding algorithms dedicated to the assessment of transformation models, drift reduction and correction light-related artifacts.

Because the presented overview is quite dense we provide a graphical summary to facilitate the comprehension of the related work and let the reader to grasp the important information quickly. The compact form the previous work is given in two Tables 3.1 and 3.2. We, however, restrict this to the mosaicing and multi-modal registration methods applied in retinal imaging only. This is because remaining image registration approaches related to other applications are secondary and the methods on assessment of transformation models, drift reduction and correction light-related artifacts can be well comprehended following the corresponding paragraphs in the text.

		Feature space	Transformation	Similarity metric	Search Strategy
Other retinal image modalities	[Pham and Abdollahi, 1991]	vessel skeleton segments	rigid	normal distance	closed-form solution
	[Mahurkar et al., 1996]	GT control points	quadratic	least squares	closed-form solution (LLS)
	[Can et al., 2002]	vessel branch + crossover points	hierarchy	sum of weighted squared distances	optimization in n-D space
	[Stewart et al., 2003]	vessel branch + crossover points	hierarchy	CEM, transfer error covariance, Beaton-Tukey bi-weight loss	extended iterative closest point algorithm
	[Yang and Stewart, 2004]	vessel branch + crossover points	quadratic	transfer error covariance + squared Mahalanobis distance	closed-form solution (joint weighted LLS)
	[Choe et al., 2006]	vessel Y-features	affine	normalized sum of squared Euclidean distances	optimization in n-D space
	[Cattin et al., 2006]	SURF	quadratic	sum of squared re-projection errors	closed-form solution (LLS)
	[Aguilar et al., 2007]	vessel branch + crossover points	quadratic	least squares	closed-form solution (LLS)
	[Li et al., 2008a]	m-space SIFT	quadratic	Sampson error	closed-form solution (LLS)
	[Lee et al., 2008]	vessel branch + crossover points	RADIC	CEM	Powell's conjugate direction method on correspondence graph
	[Wang et al., 2010]	SIFT	hierarchy	least squares	closed-form solution (NLLS)
	[Estrada et al., 2011]	all pixels, Gabor filter response features	affine	L1-norm	optimization in n-D space
	[Adal et al., 2014]	all pixels, normalized images	hierarchy	vasculature-weighted MSD	optimization in n-D space
	[Zheng et al., 2014]	vessel branch + crossover points	quadratic	to-centroid deviation	optimization in n-D space
	[Köhler et al., 2016]	all pixels	hierarchy	CC	optimization in n-D space
SLIM	[Asmuth et al., 2001]	all pixels	translation	SSD	optimization in n-D space
	[Richa et al., 2014]	SURF + all pixels in a template	rigid	SSD+local illumination compensation	optimization in n-D space
	[Zanet et al., 2016]	SURF	translation	Mahalanobis distance	optimization in n-D space

Table 3.1: Summary of the retinal mosaicing methods. n-D is a multidimensional parameter space.

	Feature space	**Transformation**	**Similarity metric**	**Search Strategy**
[Matsopoulos et al., 1999]	binary vessel maps, all pixels	bilinear	correlation	optimization in n-D space
[Choe and Cohen, 2005]	vessel Y-features	affine	MI	optimization in n-D space
[Chen et al., 2010]	PIIFD	affine	NN distance	optimization in n-D space
[Broehan et al., 2011]	vessel centerline points	quadratic	Euclidean distance	optimization in n-D space
[Ghassabi et al., 2013]	UR-SIFT and PIIFD	quadratic	leas squares	closed-for solution (LLS)
[Hernandez et al., 2015]	salient line structures	hierarchy + TPS	Champfer distance	optimization in n-D space

Table 3.2: Summary of the multi-modal retinal registration methods. n-D is a multidimensional parameter space.

Chapter 4

A Comparative Study of Transformation Models

In this chapter we present our first contribution - a comparative study of transformation models for the sequential mosaicing of long retinal sequences of slit-lamp images obtained in a closed-loop motion. First we state the motivation that has driven this study in §4.1. This is followed by the description of the imaging set-up in SLIM and the geometric assumptions we derived from it in §4.2. We provide a detailed explanation of our new efficient point correspondence based evaluation framework and error metric to compute the amount of drift in §4.3. We evaluate multiple models from existing works on retina image mosaicing as well as the homography and the TPS. We independently investigate the effects of model complexity and the number of point correspondences on drift accumulation. Finally, the results are presented and discussed in §4.4 and the conclusion is given in §4.5.

Contents

4.1 Motivation **50**
4.2 Slit-lamp imaging and geometric assumptions **50**
4.3 Transformation models and evaluation framework **50**
4.3.1 Data acquisition 51
4.3.2 Selection of pairwise point correspondences 51
4.3.3 Transformation parameter estimation 52
4.3.4 Evaluation 54
4.4 Experimental results and discussion **54**
4.4.1 Part I: effect of model complexity 55
4.4.2 Part II: effect of the number of points 57
4.5 Conclusion **58**

4.1 Motivation

As we have seen in Chapter 3, the majority of existing works on retinal image registration and mosaicing uses images obtained from a fundus camera. The quality of this type of image is higher compared to SLI. They have fewer specular reflections, good contrast and almost no blur. The transformation models applied in these works include translation, rigid motion (translation and rotation), similarity, affine and quadratic. On the other hand, mosaicing of the SLI data was covered in much fewer works mainly due the practical interest. This type of data is degraded by an uneven illumination which comes from outside the eye, especially from the contact lens. It creates viewpoint dependent artifacts, glare and specular reflections. The mosaicing method currently employed in TrackScan was presented in [Richa et al., 2014]. It demonstrates the use of the rigid transformation model. It is evident that a model of higher complexity is required due to the registration drift that often degrades retinal mosaics built with this method. However, it is not clear how far the complexity of the underlying transformation shall be extended to achieve the desired improvement in registration accuracy. Despite the variety of works which report on different transformation models for retinal image registration, only a few address their comparison and evaluation. These works, however, do not consider the mosaicing of long image sequences obtained in a closed loop motion which is typical of retinal examination with the slit-lamp. Thus, they do not address the problem of accumulated registration errors and drift. In this chapter we aim to fill this gap and find out the most suitable geometric transformation model that we can further rely on in SLIM.

4.2 Slit-lamp imaging and geometric assumptions

We used image sequences of retinal examination performed on volunteers in the University Hospital of Saint-Etienne, France. The navigated PRP system developed at QuantelMedical was used. The images were captured with a CCD camera at 60fps. Typical videos are between 2-3 minutes long. The retina is illuminated with a narrow light beam focused using a direct contact lens. The standard way of retinal examination is to perform a closed loop motion starting from the optic nerve, moving to the periphery and coming back. The camera is fixed on the moving base controlled by the ophthalmologist and undergoes translation only. Small rotations caused by head tilts occasionally occur. The spherical curvature of the retina has relatively low depth variation. The system's optics include several parts moving independently, namely the contact lens and the camera. Therefore, *the imaging device cannot be calibrated* (the relationship between a pixel's position in an image and the corresponding line of sight varies in time). Thus, there is no simple physically valid transformation to relate the images geometrically. This makes mosaicing tremendously difficult.

4.3 Transformation models and evaluation framework

Previous works do not conclude on which model best can approximate the image transformation in retinal image mosaicing. Thus, we have specifically chosen to evaluate the following seven transformation models: **T** - translation, as an intuitive choice reflecting the lateral motion of the camera; **RG** - rigid, is currently integrated in the mosaicing algorithm used in the slit-lamp device of QuantelMedical; **SM** - similarity and **AF** - affine models were chosen to check whether the modeling of slight eye movements during procedure improves accuracy; **H** - homography, as the widely used model in mosaicing [Hartley and Zisserman, 2003, Szeliski, 2006]; **QD** - quadratic, as a popular choice in retinal image registration [Can et al., 2002]; and finally the **TPS** - Thin-Plate Spline with adaptive parameter smoothing [Bartoli, 2008] which might have a great potential of success due to its elastic properties. The properties of these models are summarized in Table 4.1.

	T	**RG**	**SM**	**AF**	**H**	**QD**	**TPS**
DoF	2	3	4	6	8	12	$2k$
Linear w.r.t. source points	yes	yes	yes	yes	no	no	no
Linear w.r.t. parameters	yes	no	yes	yes	no	yes	yes

Table 4.1: Summary of the transformation models' characteristics. The DOF define the number of estimated parameters. We label each model according to whether it is linear w.r.t. its parameters or the source point. k indicates the number of control points of the TPS.

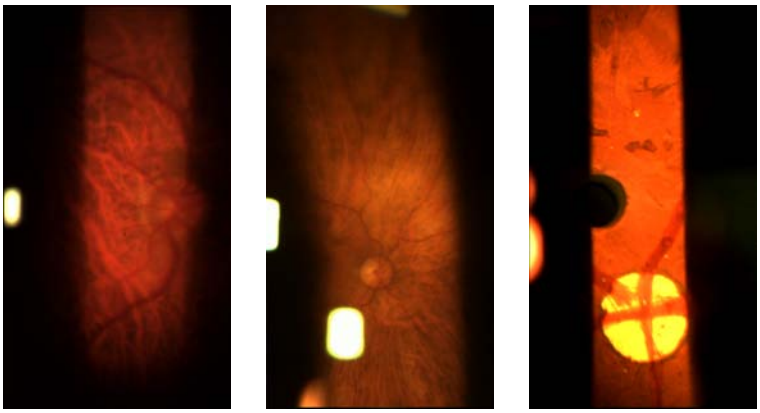
To evaluate the models' accumulated drift we propose a point correspondence based framework. The principle is to provide a noisy but outlier-free set of correspondences to minimize the effect of the fitting algorithm and evaluate the drift with an independent set of points transferred through a closed loop motion. We evaluate pairwise fitting and quantify how the model is able to connect the last and first frames in long-term image registration without using the closed loop constraint. Our framework consists of four main steps: *(1)* data acquisition and processing, *(2)* point correspondence selection, *(3)* transformation parameter estimation and *(4)* model accuracy evaluation through a number of tests. The details of each step are given further.

4.3.1 Data acquisition

Three datasets were used in this study. Sample images from each are shown in Figure 4.1. Each dataset consists of an image sequence obtained from a retinal examination video where every 5th frame was taken to ensure that each image sequence contains at least 100 frames which is at least 10 times larger compared to the mosaicing of the CF images. Two datasets were obtained from retinal examination videos of patients in the hospital. These are the ones shown in Figures 4.1a and 4.1b. One dataset of a phantom eye was included as a simplified case where the phantom was fixed on a holder and the procedure did not involve a contact lens as shown in Figure 4.1a. The length of the datasets is 254, 242 and 326 images respectively. The image size is 720×1280 pixels. The numbers of point correspondences for each dataset were not the same, resulting in as minimum 100 points per pair of frames and as maximum 400. Frames containing the minimum number of points were mostly on the periphery of the retina while frames containing more points were closer to the optic nerve. The size of the illumination slit was fixed according to the patients' comfort for the first two datasets. The visible part of the retina excluding regions of strong specularities covers at least 50% of the image.

4.3.2 Selection of pairwise point correspondences

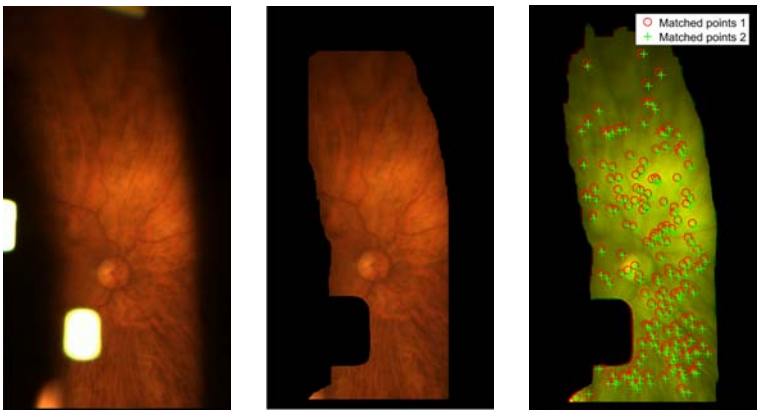
We segment the visible part of the retina and filter out strong specularities using intensity thresholding and morphological operations [Richa et al., 2014]. We then detect and extract key-points with SIFT and match them between consecutive frames. Matching is performed by measuring the L_2 norm of the difference between key-point descriptors within a pair of frames, and the basic matching algorithm suggested by [Lowe, 2004] to reject matches that are too ambiguous. A combination of automated and manual refinement steps are incorporated to exclude the remaining outliers. First we use a threshold on the points' relative displacement. The threshold is defined by summing the median and the median of absolute deviation of the points' displacement. Points which moved more than the computed threshold are discarded. Second, the manual checkup is performed with every set of point correspondences visualized on the associated images. The position of erroneous points is adjusted manually using a specifically developed GUI in Matlab. Thus, each dataset con-



(a) dataset #1 (b) dataset #2 (c) dataset #3

Figure 4.1: Sample image from each dataset.

tains between 100 and 400 correspondences $\mathbf{p} \longleftrightarrow \mathbf{q}$ obtained from f frame pairs in a closed loop $I_1 \longleftrightarrow I_2 \longleftrightarrow \ldots \longleftrightarrow I_{f-1} \longleftrightarrow I_f \longleftrightarrow I_1$. Figure 4.2 shows an example of the selection procedure where a resulting set of point correspondences between the input image and the next image in the sequence, subsampled for legibility, is shown in Figure 4.2c. Here the matched points 1 are those detected on an input image are shown with red circles and corresponding points from the next frame are shown with green crosses.



(a) Input (b) Segmentation (c) Correspondences

Figure 4.2: Selection of point correspondences.

4.3.3 Transformation parameter estimation

General Points A transformation function has the form $w(\mathbf{p}, \theta)$ where θ is a vector of transformation parameters. All transformations are estimated by minimizing the sum of squared transfer discrepancies. Due to the numerical instability of models containing cross terms and/or squared terms, as the homography and the quadratic models, the estimates might not be stable. This may be improved by data normalization, which has been well-studied for the homography [Hartley and Zisserman, 2003] but not for the quadratic model. Typical image points may have various orders of magnitude. Their increase in squared and cross terms may cause the pixel coordinates to become very large. Normalization converts pixel coordinates e.g. $\mathbf{p} \in [1; 1000]$ to normalized coordinates e.g. $\mathbf{p} \in [-1; 1]$. This is done by a simple affine transformation. The detailed explanation and substantiation of the nor-

malization procedure for the homography can be found in [Hartley and Zisserman, 2003] (Ch4.4.4). The question is whether it is possible to normalize the quadratic model with an affine transform; the answer is *yes*.

Quadratic model normalization rules We derive normalization w.r.t. the rules of function compositions. Let $N(\mathbf{p}) = \mathtt{S}\mathbf{p} + \mathbf{c}$, where $\mathtt{S} \in \mathbb{R}^{2\times 2}, \mathbf{c} \in \mathbb{R}^2$ be the normalization transform applied to the point correspondences from two consecutive frames. Let $\tilde{Q}$ be the quadratic model estimated from normalized data, using [Lawson and Hanson, 1974] for instance. Thus, to compute Q, the quadratic transform in pixel coordinates, we write:

$$Q(\mathbf{p}) = (D' \circ \tilde{Q} \circ N)(\mathbf{p}) = D'(\tilde{Q}(N(\mathbf{p}))) \tag{4.1}$$

with $D'(\mathbf{p}) = \mathtt{E}\mathbf{p} + \mathbf{k}$, where $\mathtt{E} \in \mathbb{R}^{2\times 2}, \mathbf{k} \in \mathbb{R}^2$ is the denormalization transform from the second image such that $N' \circ D' = D' \circ N' = I$.

The quadratic model is the second order Taylor series expansion of the general transformation [Can et al., 2002]:

$$Q(\mathbf{p}) = \left[\mathtt{B}_{2\times 3} | \mathtt{A}_{2\times 2} | \mathbf{t}_{2\times 1}\right] X(\mathbf{p}) \tag{4.2}$$

where $\mathtt{B} \in \mathbb{R}^{2\times 3}$, $\mathtt{A} \in \mathbb{R}^{2\times 2}$, $\mathbf{t} \in \mathbb{R}^{2\times 1}$ are the 2nd, 1st and 0th order terms of the transformation, and $X(\mathbf{p}) = [x^2, xy, y^2, x, y, 1]^\top$. We define a symmetric matrix $\hat{\mathtt{B}}_x \in \mathbb{R}^{2\times 2}$ to represent the quadratic and cross terms as:

$$\mu(\mathbf{b}_x) \stackrel{\text{def}}{=} \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} \\ \frac{1}{2}b_{12} & b_{13} \end{bmatrix} = \hat{\mathtt{B}}_x, \quad \nu(\hat{\mathtt{B}}_x) \stackrel{\text{def}}{=} \begin{bmatrix} \hat{b}_{11} \\ 2\hat{b}_{12} \\ \hat{b}_{22} \end{bmatrix} = \mathbf{b}_x \tag{4.3}$$

where $\mathbf{b}_x^\top \in \mathbb{R}^{1\times 3}$ is the first row of $\mathtt{B}$, $\mu(\mathbf{b}_x)$ is the 'packing' vector to matrix form and $\nu(\hat{\mathtt{B}}_x)$ its 'unpacking'. This is a simple reorganization of model's entries. Thus, with $\mu \circ \nu = id$ and $\nu \circ \mu = id$, we have:

$$\nu(\hat{\mathtt{B}}_x)^\top \begin{bmatrix} x^2 \\ xy \\ y^2 \end{bmatrix} = \mathbf{p}^\top \hat{\mathtt{B}}_x \mathbf{p}, \quad \mathbf{b}_x^\top \begin{bmatrix} x^2 \\ xy \\ y^2 \end{bmatrix} = \mathbf{p}^\top \mu(\mathbf{b}_x)\mathbf{p} \tag{4.4}$$

Each dimension of $\tilde{Q}$ can then be written as:

$$\tilde{Q}_x(\mathbf{p}) = \mathbf{p}^\top \hat{\mathtt{B}}_x \mathbf{p} + \mathbf{a}_x^\top \mathbf{p} + t_x \tag{4.5}$$

where $\mathbf{a}_x^\top \in \mathbb{R}^{1\times 2}$ is the first row of $\mathtt{A}$ and t_x is the first element of $\mathbf{t}$.

First, to compose the quadratic model with a normalization transform N we use composition rules expressed in (4.1) and (4.5). We write the composition as follows:

$$\begin{aligned} (\tilde{Q}_x \circ N)(\mathbf{p}) &= \frac{1}{2}(\mathtt{S}\mathbf{p} + \mathbf{c})^\top \hat{\mathtt{B}}_x (\mathtt{S}\mathbf{p} + \mathbf{c}) + \mathbf{a}_x^\top (\mathtt{S}\mathbf{p} + \mathbf{c}) + t_x \\ &= \frac{1}{2}\mathbf{p}^\top \mathtt{S}^\top \hat{\mathtt{B}}_x \mathtt{S}\mathbf{p} + (\mathbf{c}^\top \hat{\mathtt{B}}_x + \mathbf{a}_x^\top)\mathtt{S}\mathbf{p} + (\frac{1}{2}\mathbf{c}^\top \hat{\mathtt{B}}_x + \mathbf{a}_x^\top)\mathbf{c} + t_x \\ &= \left[\tfrac{1}{2}\nu(\mathtt{S}^\top \hat{\mathtt{B}}_x \mathtt{S})^\top \quad (\mathbf{c}^\top \hat{\mathtt{B}}_x + \mathbf{a}_x^\top)\mathtt{S} \quad (\tfrac{1}{2}\mathbf{c}^\top \hat{\mathtt{B}}_x + \mathbf{a}_x^\top)\mathbf{c} + t_x\right] X(\mathbf{p}) \end{aligned} \tag{4.6}$$

which shows that $\tilde{Q}_x \circ N$ is a quadratic transformation which follows that $\tilde{Q}_y \circ N$ is a quadratic transformation too.

To compose the denormalization transform D' with the quadratic model resulting from (4.6), we follow the previous derivation and write the composition as follows:

$$(D' \circ Q)(\mathbf{p}) = \mathtt{E}(\left[\mathtt{B}|\mathtt{A}|\mathbf{t}\right] X(\mathbf{p})) + \mathbf{k} = \left[\mathtt{EB}|\mathtt{EA}|\mathtt{E}\mathbf{t} + \mathbf{k}\right] X(\mathbf{p}) \tag{4.7}$$

which shows that $D' \circ Q$ is a quadratic transformation. Consequently, this establishes that, $D' \circ \tilde{Q} \circ N$ is a quadratic transformation too and that normalized estimation of the quadratic transformation is possible.

Normalized estimation of the quadratic transformation The following steps summarize the normalization procedure and parameter estimation for the quadratic model:

1. **Normalize:** define N and D' from the image size and normalize the point correspondences $\mathbf{p} \longleftrightarrow \mathbf{q}$ from two consecutive images with N to obtain $\tilde{\mathbf{p}} \longleftrightarrow \tilde{\mathbf{q}}$
2. **Fit $\tilde{Q}$:** apply the LLS algorithm to the normalized point correspondences $\tilde{\mathbf{p}} \longleftrightarrow \tilde{\mathbf{q}}$ to obtain $\tilde{Q}$.
3. **Find** use equations (4.6) and (4.7) to get the final Q.

We denote the normalized quadratic model as **QDn** and include it for evaluation. The effect of this normalization is also discussed and illustrated in §4.4.1.

4.3.4 Evaluation

To independently evaluate the effect of the model complexity and the number of point correspondences we analyze two types of error metrics as illustrated in Figure 4.3. We compute the Local Fitting Error (LFE) - the discrepancy of data point and corresponding model estimate in pixels. This allows us to evaluate model fitting in pairwise registration as follows:

$$\xi_{LFE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \parallel \mathbf{q}_i - w(\mathbf{p}_i, \theta) \parallel^2} \tag{4.8}$$

where $\mathbf{p}_i \longleftrightarrow \mathbf{q}_i, i = 1, ..., n$ are all the point correspondences.

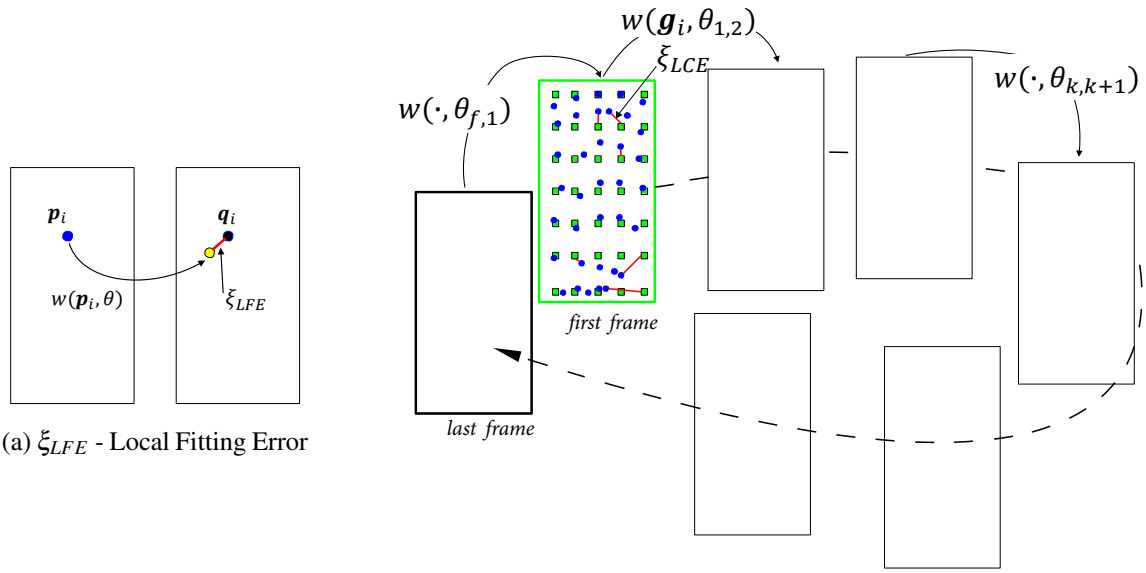
We propose a Loop Closure Error (LCE) metric. This shows how the composition of estimated transformations affects the global registration and accumulated drift. The idea is to initialize a uniform grid of points $g_1, ..., g_l$ at the first frame of the sequence and use the set of pairwise estimated transformations applied sequentially to transfer the grid throughout the sequence. The discrepancy between the initial and resulting sets of points is then measured in pixels as follows:

$$\xi_{LCE} = \sqrt{\frac{1}{l}\sum_{i=1}^{l} \parallel g_i - \zeta_i \parallel_2^2} \tag{4.9}$$

where $\zeta_i = w(...(w(g_i, \theta_{1,2}))..., \theta_{f,1})$.

4.4 Experimental results and discussion

Our evaluation has two parts. In Part I the ξ_{LFE} and ξ_{LCE} metrics were computed for every model on three datasets where all the pairwise point correspondences were used for parameter estimation. This is to analyze how the model complexity affects the local registration error and accumulated drift. The narrow FOV, poorly textured regions of the retina and small amount of landmarks sometimes complicates the automatic detection of a sufficient number of point correspondences. A suitable transformation model has to cope with this limitation. Thus, in Part II we study the effect of varying the number of point correspondences.



(a) ξ_{LFE} - Local Fitting Error

(b) ξ_{LCE} - Loop Closure Error

Figure 4.3: Error metrics for transformation model complexity evaluation.

4.4.1 Part I: effect of model complexity

The results given in Table 4.2 show that ξ_{LFE} decreases with increasing complexity of the model. We found that T, RG and SM provide similar results. The difference between RG and SM is negligible. This is because isotropic scaling is almost minimal in slit-lamp imaging. H, despite its complexity over AF, generally gives similar results to AF and even inferior in datasets #2 and #3. This raises the question of whether perspective matters. The answer would be *no*. Modeling perspective is not useful for curved retina and purely lateral motion of the camera. Finally, the TPS provides the smallest ξ_{LFE} in datasets #1 and #2 and QD gives the smallest ξ_{LFE} in dataset #3.

	dataset # 1		**dataset # 2**		**dataset # 3**	
	ξ_{LFE}	ξ_{LCE}	ξ_{LFE}	ξ_{LCE}	ξ_{LFE}	ξ_{LCE}
T	3.185	62.906	3.165	18.441	3.179	59.892
RG	3.164	57.262	3.066	50.288	3.161	75.162
SM	3.162	72.473	3.064	49.743	3.158	76.175
AF	3.105	102.150	2.986	78.785	3.056	221.050
H	3.073	201.950	3.000	333.650	3.066	351.390
TPS	3.019	125.150	2.864	275.920	2.971	191.790
QD	$F(28)$	$F(28)$	$F(149)$	$F(149)$	2.762	478.070
QDn	$F(56)$	$F(56)$	2.866	254.870	2.886	236.330

Table 4.2: Average ξ_{LFE} and ξ_{LCE} across the different datasets.

ξ_{LCE}, in contrast, shows a performance superior to simpler models. RG gives the smallest ξ_{LCE} for dataset #1, while T is best in datasets #2 and #3. Following the same pattern as for ξ_{LFE}, RG and SM have errors with difference close to 1 pixel for datasets #2 and #3. However this does not hold for dataset #1. As one can see the difference in ξ_{LCE} between T and RG in dataset #1 is small (only 5.644 pixels) while for datasets #2 and #3 it is much larger (31.847 and 15.27 pixels respectively). This indicates that the rotation component of the model was completely redundant when the patient froze during examination (dataset #2) and the phantom eye was fixed to the holder (dataset #3). AF

and TPS showed close results in datasets #1 and #3 while for dataset #2 ξ_{LCE} differs considerably. Additionally, ξ_{LFE} was similar between AF and TPS for dataset #2. This means a small impact of affine deformations in datasets #1 and #3. H appeared to be the worst model.

QD was derived specifically to fit the curved retina [Can et al., 2002]. However, it turned out that its estimation from our data is not stable. As one can see this model gives the smallest ξ_{LFE} for dataset #3 only. This, somehow, correlates with results described in the related literature [Adal et al., 2014, Zheng et al., 2014, Stewart et al., 2003, Can et al., 2002]. However, this model completely fails in ξ_{LCE} as indicated with $F(x)$ where x is a number of the frame where failure occurred. Indeed, the accumulated drift causes some models to prematurely stop registration before the end of the sequence. In such cases, the model contains numerically unstable parameter combinations (quadratic and cross terms) which force point coordinates to become very large if a 'faulty' estimate occurs in the process of chaining for ξ_{LCE} computation. Therefore, when the points tend to be in a degenerate configuration it is the most sensitive model. Thus, we rule out QD from the next experiment. Our normalization method improves the fitting of the quadratic model. Results for dataset #1 showed that failure has been delayed by QDn for 28 frames. The failure was completely eliminated in datasets #2 and #3. We illustrate this improvement with graph plots in Figure 4.4.

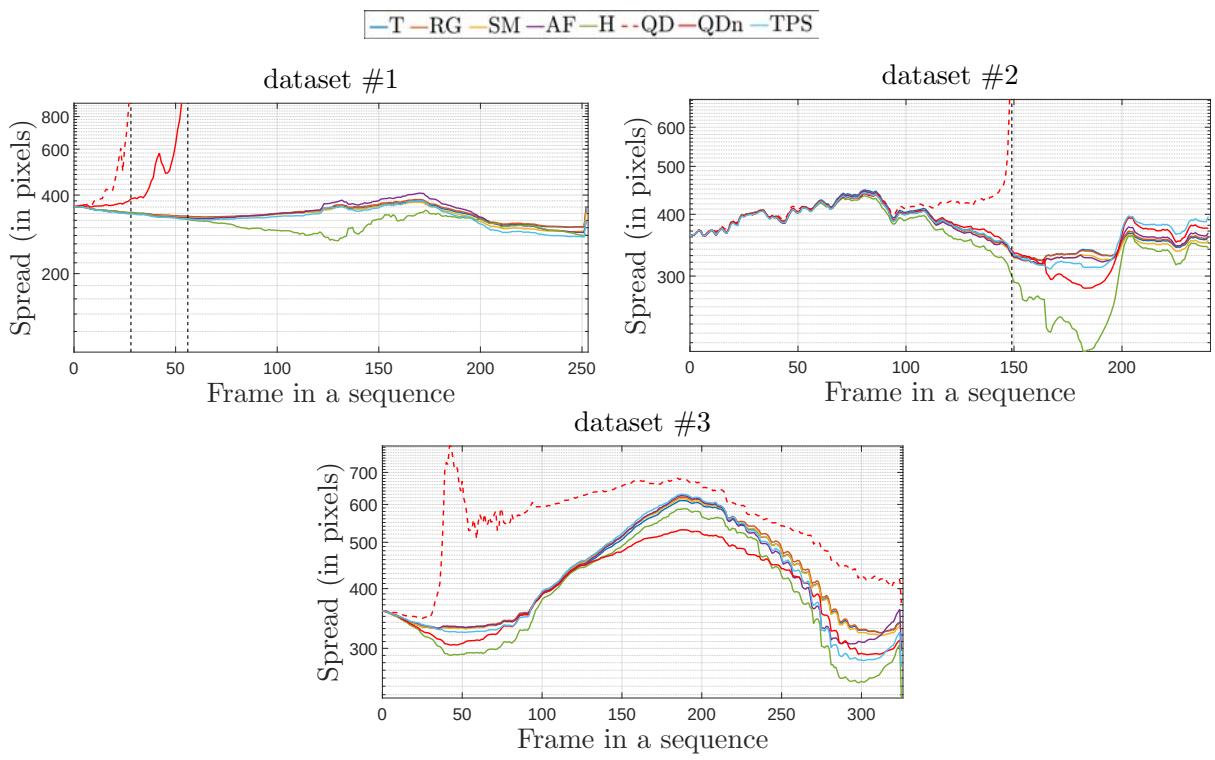
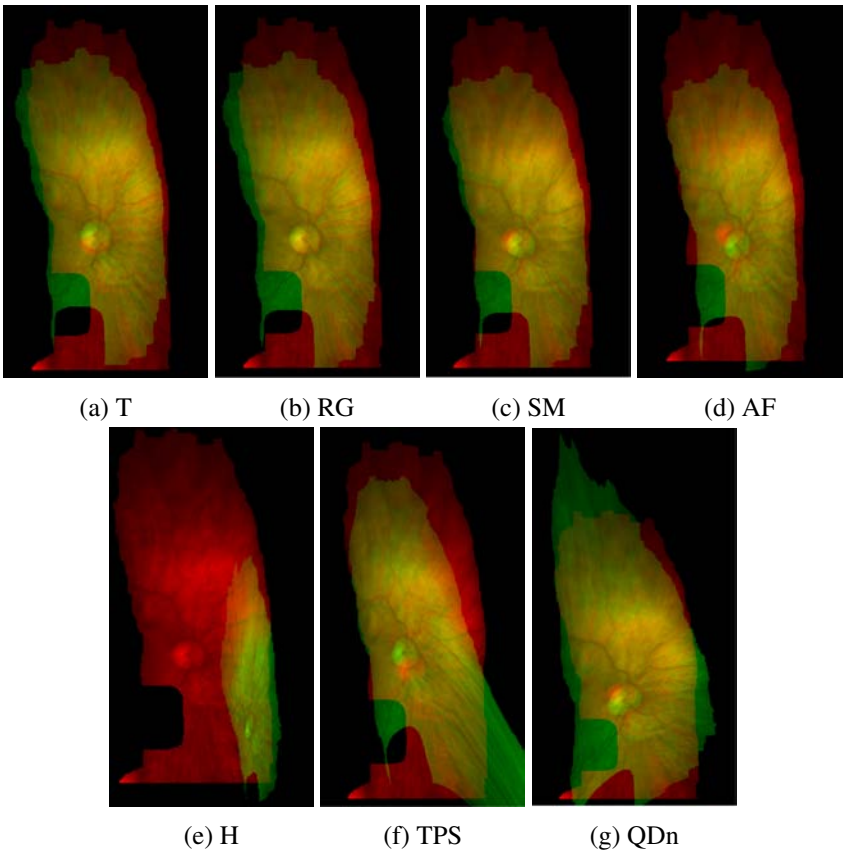


Figure 4.4: The 'spread' evaluation results over different datasets without subsampling.

We show the 'spread' of the points from the uniform grid defined for ξ_{LCE} computation. This demonstrates the model response to scene geometry at central and peripheral portions of the retina. The dashed black lines indicate the frame when failure occured. One can observe that QDn provides the smallest ξ_{LFE} for dataset #3 and nearly the same ξ_{LFE} as TPS for dataset #2. One can see that normalization suppressed the effect of the quadratic part making QDn fit similar to AF in dataset #3. Examples of registered image pairs highlighting areas in which the output of the evaluated transformation models differ are shown in Figure 4.5.



(a) T (b) RG (c) SM (d) AF

(e) H (f) TPS (g) QDn

Figure 4.5: Examples of registered image pairs with different transformation models. The images are taken from dataset #2. The first image of the sequence is registered with the last image by applying the set of 241 pairwise estimated transformations sequentially.

4.4.2 Part II: effect of the number of points

We defined the minimum and maximum number of samples as 20 and 100 respectively and computed the ξ_{LFE} and ξ_{LCE} by selecting points randomly with steps of 2 samples. We made 50 trials and averaged the results. Results for this evaluation test are similar among the three datasets. The example of dataset #2 is shown in Figure 4.6. All transformations show an increase in ξ_{LFE} approximately 0.5 pixels with an increase in the number of point correspondences. This happens because more data brings more constraints to the estimated parameters. However, there is no common trend among results on ξ_{LCE}. Varying subsets of points from 20 to 100 lead ξ_{LCE} to decrease approximately 1.5 times for T, RG, SM and AF. It also decreased approximately 2.5 times for H. One can see that H shows high variance when the number of points is not sufficient and stabilizes only when more than 50 points are supplied. TPS showed a decreasing trend between 20 and 45 points followed by unstable behavior in 45-78 points and starts increasing between 78 to 100 points. This instability is due to the number of control points used to define the deformation grid in TPS, it was constant despite of changing the number of point correspondences. QDn started to give meaningful results only when 68 points were supplied for estimation. It showed an unstable behavior with varying ξ_{LCE} from 256 to 263 between 70 and 100 points. This indicates that this model is very sensitive to the number of points.

The results obtained on three datasets have shown that local registration error decreases with increasing complexity of the transformation model while simple models appeared to produce less accumulated drift. The homography turned out to be irrelevant as perspective deformations might be

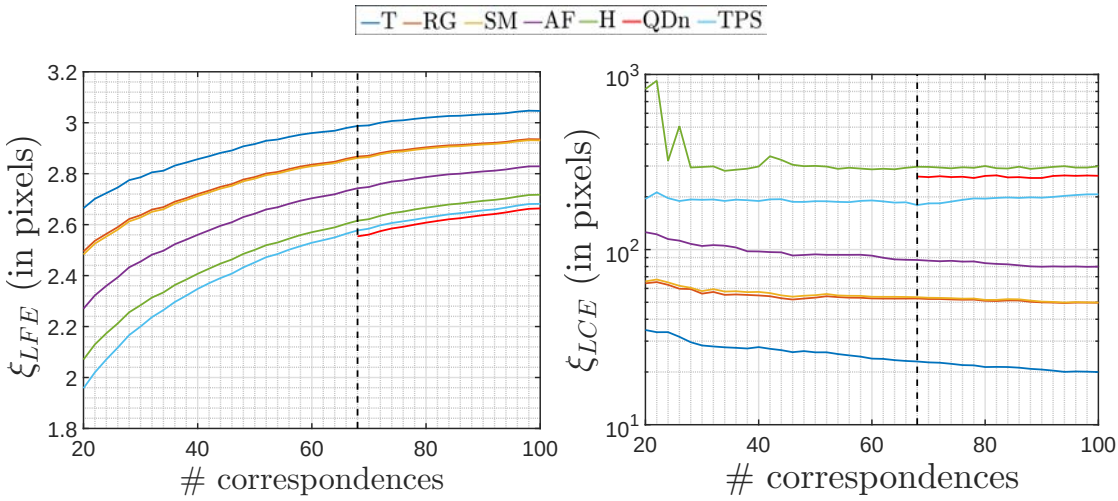


Figure 4.6: Effect of the number of points. Example of dataset #2.

considered absent. Despite its popularity in the application of retinal image registration the quadratic model turned out to be completely unstable on our data even after improvement by our proposed normalization procedure. Thus, the choice is meant to be done from the remaining models, namely translation, similarity, rigid, affine and TPS.

The translation, despite of the lowest accumulated drift, is too simple for the majority of clinical cases where the patients are normally very photosensitive and cannot completely freeze during the procedure. The rigid transformation, which is currently used in the TrackScan and can model rotations, is not sufficiently flexible. The similarity covers isotropic scaling which sometimes occur during the examination. The affine model represents a superset for translation, rigid and similarity models. It covers more deformation types and provides better results. The TPS is complex but the adaptive smoothing makes it always stiff causing, however, a large drift. Therefore, in sequential mosaicing with long slit-lamp image sequences the simple models, specifically translation, rigid, similarity and affine can be the choice among others. However, an affine model is the best possible compromise between ability to model pairwise transformation and simplicity in dealing with drift. The models with higher complexity are best for short-term registration on different types of data.

4.5 Conclusion

In this chapter we have presented a comparative study of transformation models applied to sequential retinal image mosaicing in computer-assisted slit-lamp imaging. We proposed the point correspondence based evaluation framework to assess different geometric transformation models on the subject of drift accumulation. This led us to conclude that the affine transformation is the most suitable model for SLIM.

Chapter 5

Drift Reduction

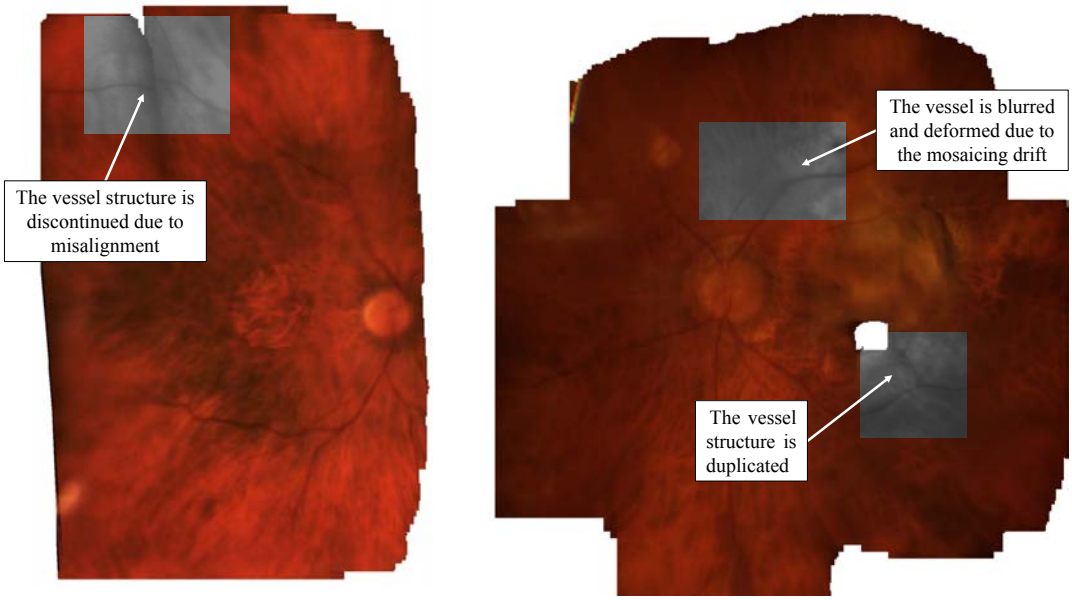
In this chapter we present our second contribution - a method for drift reduction specifically designed for the case of long-image sequences in SLIM. We start by explaining the motivation that has driven this study in §5.1. We provide a detailed description of the proposed drift reduction method in §5.2. Our main idea is to create long-term high precision point correspondences by associating a simple global model with local correction and perform key-frame based Bundle Adjustment. In this section we also introduce a new measure for accumulated drift. Finally, the extensive evaluation and comparative results with the state-of-the-art method in SLIM that show significantly lower accumulated drift are presented in §5.3. The summary follows in §5.4.

Contents

5.1 Motivation **60**
5.2 Methodology **60**
5.2.1 Mosaicing initialization 60
5.2.2 Motion estimation 61
5.2.3 Prediction 61
5.2.4 Track correction 61
5.2.5 Key-frame instantiation and Local Bundle Adjustment 62
5.3 Experimental results and discussion **63**
5.3.1 Dataset acquisition 63
5.3.2 Evaluation 63
5.4 Conclusion **68**

5.1 Motivation

The main difficulty in SLIM is the accumulated mosaicing drift due to the small number of features away from the optic disc, the distortion induced by the geometry of the eye and the contact lens that causes illumination artifacts affecting motion estimation. A common approach to image mosaicing is to compute transformations only between temporally consecutive images in a sequence, and then use the rule of composition to obtain the transformation between non-temporally consecutive views. Many aforementioned mosaicing algorithms including the one implemented on the TrackScan platform [Richa et al., 2014] follow this approach. Despite the low computational cost and simplicity of this method, due to its 'chaining' nature, alignment errors tend to accumulate, causing images to drift in the mosaic. Examples of this can be seen in Figure 5.1. Our main motivation in this work is to reduce the drift and boost the geometric quality in SLIM.



(a) (b)

Figure 5.1: Examples of mosaics obtained with [Richa et al., 2014]. *(a)* - registration drift is visible through the mismatched vascular structure, *(b)* - example with drift induced blurred regions and duplication. The visual assessment was performed by an expert.

5.2 Methodology

Our algorithm consists of the following steps: *(1) Initialization* with a key-point detector, *(2) Motion Estimation* based on key-frames selection, *(3) Prediction*, using a popular tracking algorithm, *(4) Track Correction* using a simple global model with local adjustment *(5) Key-frame Instantiation and Local BA*. The detailed description is given in the subsequent paragraphs.

5.2.1 Mosaicing initialization

We use latin bold to refer to key-points (such as $\mathbf{p}$) and Greek chatacter τ to refer to point tracks. Frame indexing is denoted as $f = 1, ..., n_f$ and $k = 1, ..., n_k$ is used for key-frame indexing. We start by obtaining a set of key-points $\{\mathbf{p}_i\}_{i=1}^{n_i}$ detected on the first frame $I_{f=1}$ and defining an initial set of tracks $\{\tau_j\}_{j=1}^{n_j} = \{\mathbf{p}_i\}_{i=1}^{n_i}$. We also tag the first frame as a key-frame $I_{f=1} \rightarrow I_{k=1}$. Here and in the

following steps all the computation and processing is performed on the image where only the visible part of the retina is kept and strong specular reflections have been filtered out. A segmentation mask is produced such that image pixels which do not belong to the retina are assigned to zero (*ie.*zero-intensity pixels) and to one otherwise. This is done using thresholding followed by morphological refinement [Richa et al., 2014]. In the experimental section we assess different types of key-point detectors, SIFT, the Minimum Eigen Value algorithm (minEig) [Tomasi and Kanade, 1991] and their impact on the performance of the proposed algorithm. We also use a Uniform Grid of points (UGrid) evenly placed on the area of the visible retina to complement the evaluation.

5.2.2 Motion estimation

Inter-frame motion estimation with a simple model as used in [Richa et al., 2014] seems to be robust but inaccurate, typically up to 5 pixels as was discovered in Chapter 4. We can use this simple global model to create better inter-frame correspondences, and then tracks. The slit-lamp system's optics include several parts moving independently, namely the contact lens and the camera. This complicates the derivation of an accurate, simple and physically valid transformation to relate the images geometrically. We use the affine transformation in our work as a best tradeoff as resulted from comparative study in Chapter 4. When the new frame I_f comes we estimate the motion to the last key-frame $\mathbf{A}_{f\rightarrow k-1}$ by solving the LLS problem where we minimize the sum of squared transfer discrepancies:

$$\tilde{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n_i} \| \mathbf{q}_i - w(\mathbf{p}_i;\theta) \|_2^2 \tag{5.1}$$

where $\tilde{\theta}$ is an estimated (6×1) vector of motion parameters of the last key-frame. The transformation function has the form $w(\mathbf{p};\theta)$ and $\mathbf{p}_i$, $\mathbf{q}_i$ are key-point correspondences between the current and previous frames.

5.2.3 Prediction

We propagate the existing *query* tracks τ_j using popular tracking algorithm Kanade-Lucas-Tomasi (KLT) [Shi and Tomasi, 1994] to obtain the *candidate* tracks as:

$$\tau'_j = KLT(\tau_j, I_{f-1}, I_f) \tag{5.2}$$

The key-point associated with the *candidate* track is then checked for zero-intensity (*ie.*intensity values of all color channels equal to zero). If true it is then rejected as a faulty prediction because the track is considered valid only if it belongs to the visible part of the retina. We have chosen KLT as it is an appearance based method which uses local search. It is fast and robust just enough to handle changes between consecutive frames. It can cope with sudden motion better compared to statistical approaches such as the Extended Kalman Filter (EKF) where the redundancy exists in time.

5.2.4 Track correction

We proceed with the refinement procedure to correct the position of the predicted *candidates* as schematically illustrated in Figure 5.2. We first warp the new image using the previously estimated affine transformation as:

$$I_f^{\omega} = \omega(I_f, \mathbf{A}_{f\rightarrow k-1}) \tag{5.3}$$

We perform an exhaustive search in a 5×5 neighborhood w around the *query* tracks locations on the warped image I_f^{ω} to find a possible update $\tilde{\tau}_j$ by minimizing a similarity metric. We search on the warped image because it allows us to find an estimate in a local area which is directly related to the perceived misalignment. We evaluate several metrics in this study, namely the SSD, NCC and

Sum of Hamming Distances (SHD). Finally the corrected position of the predicted track locations is computed using the previously estimated motion as:

$$\bar{\tau}_j = \phi(\tilde{\tau}_j, \mathbf{A}_{f\to k-1}) \tag{5.4}$$

where ϕ is the back-warping function.

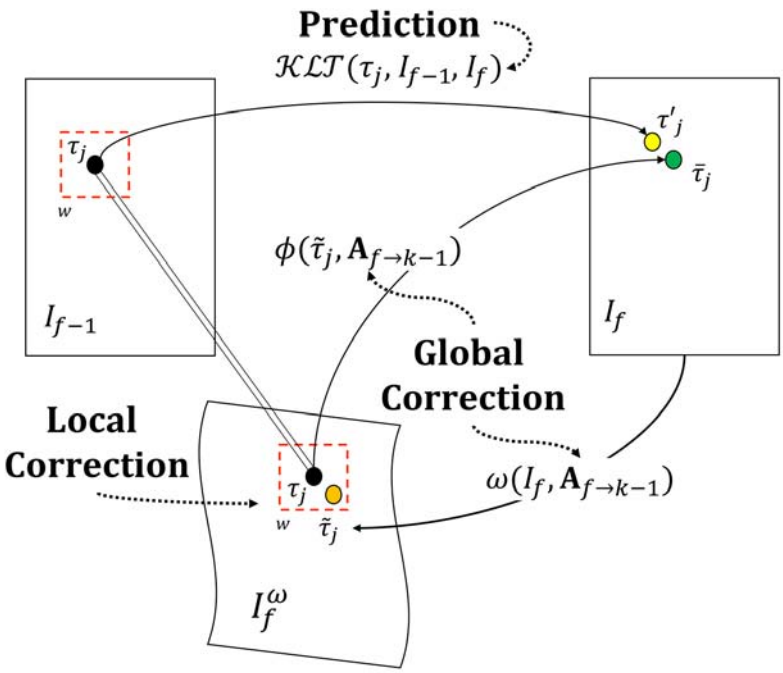


Figure 5.2: Schematic illustration of track prediction and correction on a sample track τ_j.

5.2.5 Key-frame instantiation and Local Bundle Adjustment

We compute the tracking loss L in the current frame as the percentage of lost tracks from the last key-frame to provide the condition for inclusion of new key-points and then tracks $L = \frac{sizeof(\tau \in I_f)100}{sizeof(\tau \in I_k)}$. This does not indicate re-initialization of the tracking process in case of full occlusion. It rather allows us to assure that sufficiently many points are tracked at all times. Thus, if $L > 50\%$, we detect new key-points τ_f^{new} as in the *Initialization* step. We then filter out those new tracks which fall in the predefined local neighborhood (7×7 pixels in our experiments) and join the two sets of tracks. This is done to keep new tracks not too close to the existing ones and avoid populating new tracks with redundant locations. Finally, the current frame is tagged as new key-frame $I_f \to I_{k+1}$.

We then invoke a local BA-type routine. The idea is to minimize the re-projection error. An unknown 2D point $\mathbf{g}_j$ is associated with each track $\tau_{k,j}$ and an affine transform $w(\mathbf{g}_j;\theta)$ with each key-frame. The presence/absence of a track in a key-frame is given by an indicator variable $v_{k,j} \in \{0,1\}$. The re-projection error to minimize is:

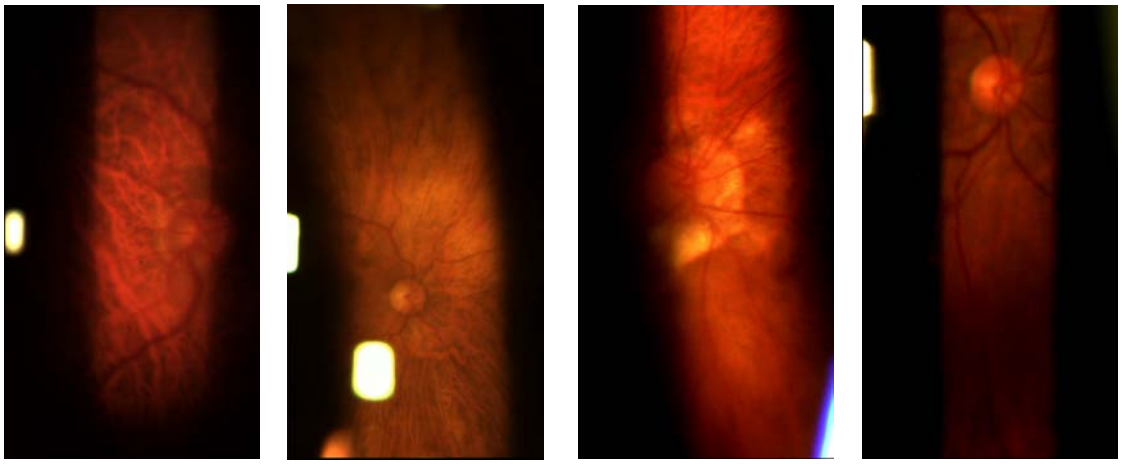
$$\underset{\mathbf{g}_j,\theta}{\operatorname{argmin}} \sum_{k=1}^{n_k} \sum_{j=1}^{n_j} v_{k,j} \parallel \tau_{k,j} - w(\mathbf{g}_j;\theta) \parallel_2^2 \tag{5.5}$$

we solve this with matrix factorization in the LLS sense [Hartley and Zisserman, 2003]. We repeat from *Motion Estimation* step for the rest of the sequence.

5.3 Experimental results and discussion

5.3.1 Dataset acquisition

The datasets used for evaluation were obtained from four retinal examination videos of volunteers in the University Hospital of Saint-Étienne, France. Figure 5.3 shows the sample images corresponding to each dataset. We took every 5th frame to produce images sequences to simplify the evaluation routine. Thus, each resulting video spans at least 100 frames. The standard way of retinal examination is to perform a closed loop motion starting from the optic nerve, moving to the periphery and coming back. Full occlusion may occur due to a patient's sudden move and/or specular reflections induced by the contact lens. Dealing explicitly with such challenging conditions is out of the scope of this work. Thus, our video samples were chosen in such a way that no full occlusion occurred in a sequence.



(a) (b) (c) (d)

Figure 5.3: Sample images from different slit-lamp datasets. *(a)* - dataset#1, 253 images, *(b)* - dataset#2, 242 images, *(c)* - dataset#3, 169 images, *(d)* - dataset#4, 309 images.

5.3.2 Evaluation

In retinal imaging it is difficult to evaluate mosaicing methods objectively due to the lack of ground-truth for alignment. A method of generating simulated retinal image sets by modeling geometric distortions and the image acquisition process have been proposed for the case of fundus images [Lee et al., 2007]. However, in slit-lamp imaging this option is not directly applicable and the adjustment of this technique to our case is out of the scope of this work. Simulation of the imaging process with a virtual camera becomes problematic likewise due to the complexity of the optical set-up. We provide objective quantitative partial performance evaluation of our method in two stages. First, the assessment of the steps of the method which potentially have strong influence on the result evaluated. This is followed by a comparison of the best performing combination to [Richa et al., 2014].

Does the metric matter?

To assess the impact of the chosen local similarity metric on the precision of the track correction we compare different metrics namely SSD, NCC and SHD. Both SSD and NCC metrics were considered as the popular choice in real-time tracking algorithms and due to the simplicity of the computation. These are correlation based metrics which rely only on the intensity information. The

SHD metric on the other hand is well known for its usage in binary feature matching. It is very fast to compute and it captures structural information, which is a favorable feature in case of slit-lamp imaging where illumination variations are often present. We compute the Forward-Backward Constancy (FBC) error. The idea is to track the $\bar{\tau}_{i,j}$ *backward* continuously performing *Prediction* and *Track Correction* steps. The FBC error is defined as the distance in pixels from the original location of the track to the final location after the backward tracking. We define the acceptance threshold as 3 pixels. Table 5.1 shows the computed FBC across datasets. We calculate FBC every time when the correction step is invoked and take an average among all measurements. We show results for different key-point detectors used to initialize the tracks. As one can see, SHD generally provides a lower error among the datasets while SSD comes second and NCC turned out to be the inferior one.

		dataset#1	**dataset#2**	**dataset#3**	**dataset#4**
UGrid	**SSD**	3.88	2.71	4.26	3.53
	NCC	4.29	3.41	5.34	5.23
	SHD	2.82	2.70	3.82	**2.36**
minEig	**SSD**	3.47	2.64	4.71	3.15
	NCC	4.82	3.70	6.97	5.27
	SHD	**2.64**	**2.05**	3.53	2.65
SIFT	**SSD**	3.56	2.70	4.21	3.95
	NCC	3.62	3.82	5.28	5.18
	SHD	2.81	2.68	**3.18**	2.89

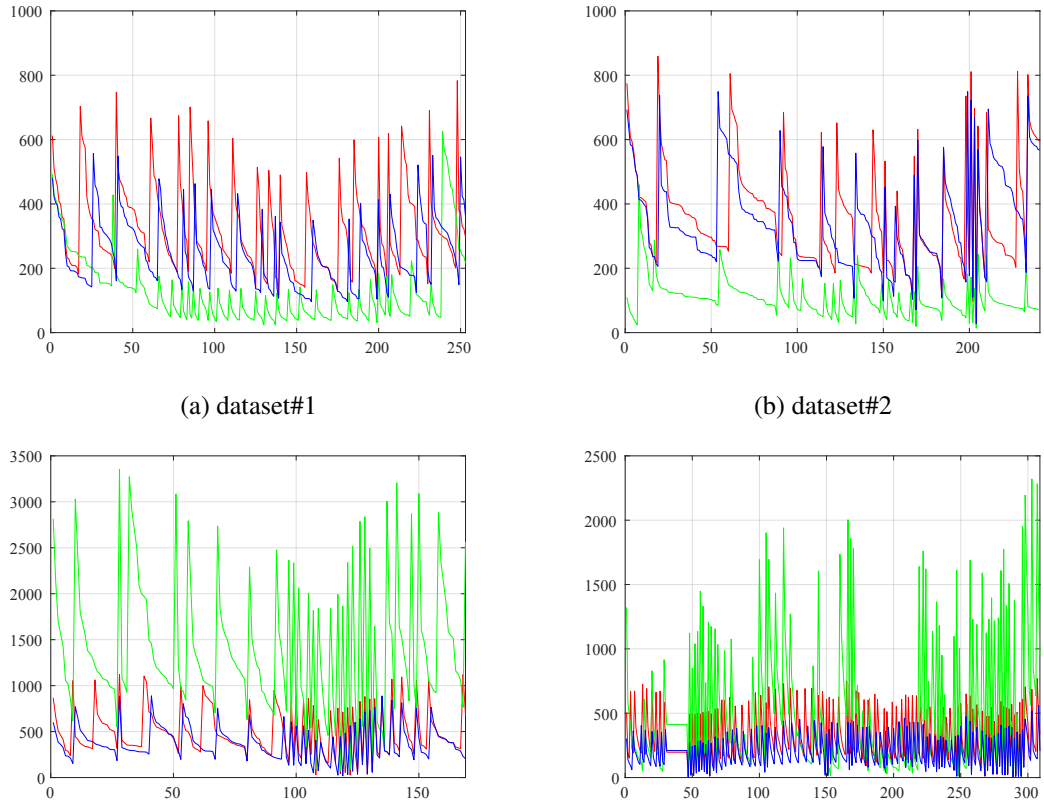
Table 5.1: Forward-Backward Consistency for similarity metrics evaluation.

How long the tracks are?

Long-term tracks is a fundamental part of BA-type refinement. Thus, the quality of the method is directly related to the average length of the tracks, the longer the better. We assess the length of the tracks with and without the correction step of our method. To evaluate this we compute the average length of the tracks across different subsets of frames which were established each time a new key-frame was defined. We call it the span, denoted S. We also check the average number of tracks per frame for a given dataset, denoted μ, as it has a heavy impact on the propagation of local alignment errors. Finally, we analyze the number of key-frames instantiated for a given dataset, denoted κ, as an additional indicator of track accuracy, the lower the better.

The graph plots given in Figure 5.4 show the number of tracks per frame for tracking without correction using three options to define the key-points. The tracks obtained with UGrid are shown as red curve, minEig is in green and SIFT was used to obtain the tracks which are shown in blue. As one can see, defining the uniform grid of points to initialize the tracks gives higher track/frame rate for the datasets #1 and #2. However, minEig produces more tracks for datasets #3 and #4. One can also observe that the second dataset seems to be an easy example due to the the lower amount of spikes presented on the graph. In fact, the spikes on the graph are the events when the new key-frame was instantiated and new tracks were added to the existing ones. Similarly, one can conclude that dataset#4 is the most difficult case for evaluation. This is not only because it has the longest sequence but also because the retina was not properly illuminated during the examination, thus, not providing sufficient reliable information.

The tracking statistics across datasets for this experiment are shown in Table 5.2. One can see that for dataset#1 the maximum span was achieved using UGrid from initialization. However, SIFT shows more consistent tracks for dataset#2. Finally, minEig appears to perform better on datasets #3 and #4. The average number of tracks per frame follows a similar behavior resulting in more tracks for datasets #1 and #2 with a uniform grid while for datasets #3 and #4 more tracks are given by



(a) dataset#1

(b) dataset#2

(c) dataset#3

(d) dataset#4

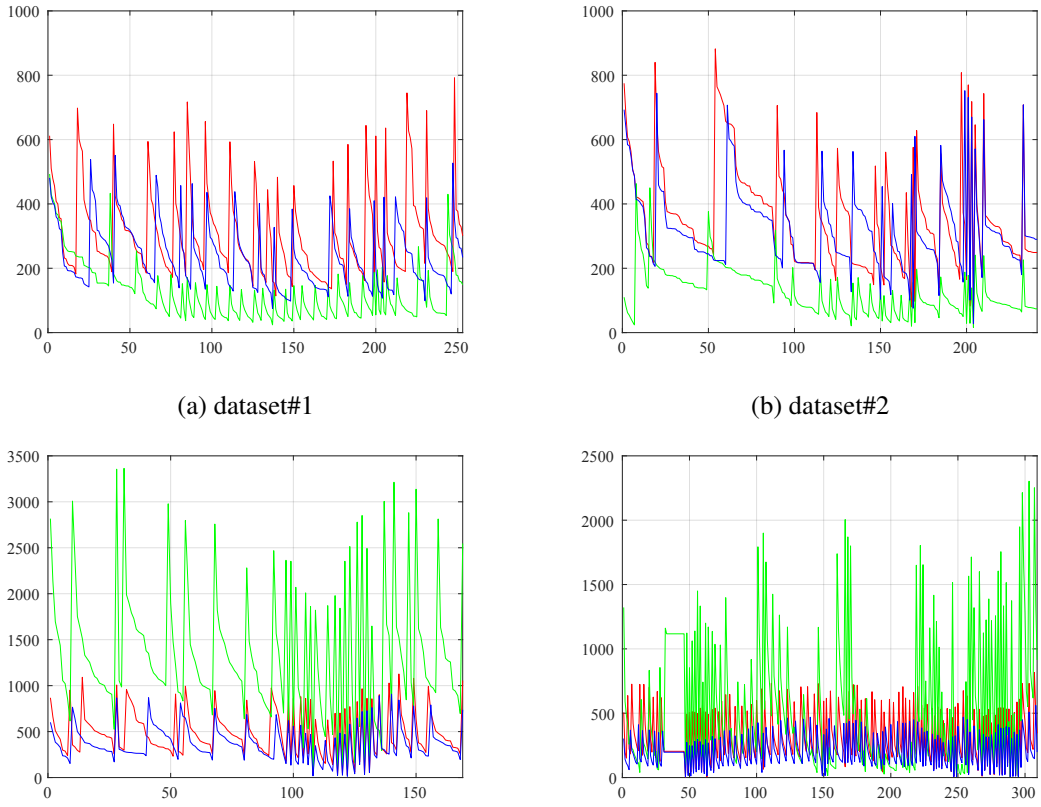
Figure 5.4: Number of tracks versus frames. Results show performance with UGrid key-points *(red)*, minEig *(green)* and SIFT key-points *(blue)* respectively on the experiment *without* track correction.

	UGrid				**minEig**				**SIFT**			
	μ	κ	S_{mean}	S_{max}	μ	κ	S_{mean}	S_{max}	μ	κ	S_{mean}	S_{max}
dataset#1	322	19	8	49	128	29	5	39	233	19	9	48
dataset#2	358	19	7	39	106	25	5	37	350	17	9	41
dataset#3	493	30	3	19	1398	30	3	20	372	28	3	18
dataset#4	321	95	1	11	525	87	2	12	205	92	2	11

Table 5.2: Tracking statistics *without* track correction. μ - average number of tracks per frame, κ - number of key-frames, S_{mean} - average span, S_{max} - maximum span.

minEig. Overall it can be concluded that initializing with UGrid seems to be a tradeoff when we do not incorporate track correction.

What happens once the correction step is included in the method? The results of this setting are given in Figure 5.5. The graphs demonstrate that the number of tracks per frame slightly increased for all the datasets. This is supported by the statistics provided in Table 5.3. Indeed, using the result of the evaluation of the similarity metrics, namely SHD, we obtain improvement for all the statistics and for a number of key-framesκ especially. This indicates that the track correction step using a simple global model with local neighborhood based adjustment is an efficient way to obtain longer tracks with better precision.



(a) dataset#1 (b) dataset#2

(c) dataset#3 (d) dataset#4

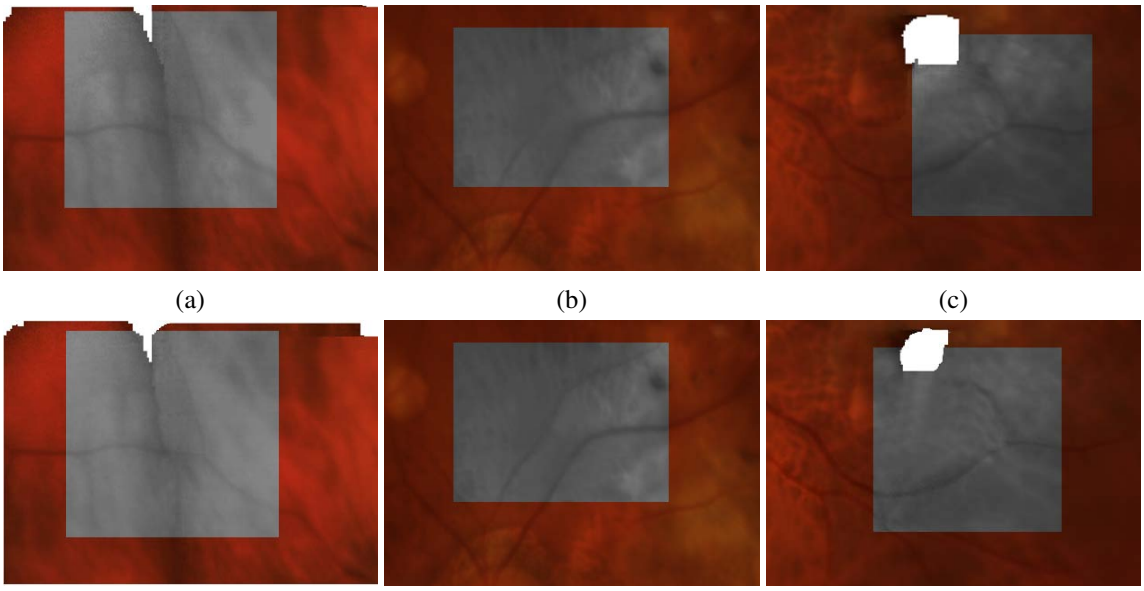
Figure 5.5: Number of tracks versus frames. Results show performance with UGrid key-points *(red)*, minEig *(green)* and SIFT key-points *(blue)* respectively on the experiment *with* track correction.

	UGrid				**minEig**				**SIFT**			
	μ	κ	S_{mean}	S_{max}	μ	κ	S_{mean}	S_{max}	μ	κ	S_{mean}	S_{max}
dataset#1	308	15	11	52	121	28	6	39	233	16	12	52
dataset#2	359	16	12	44	124	20	7	35	320	12	10	45
dataset#3	485	24	5	26	1351	26	5	24	370	27	4	21
dataset#4	320	90	2	12	550	80	4	15	205	86	3	13

Table 5.3: Tracking statistics *with* track correction. μ - average number of tracks per frame, κ - number of key-frames, S_{mean} - average span, S_{max} - maximum span.

Are we reducing the drift?

As resulted from the previous experiments, tracking initialized with uniform grid and the SHD based track correction scheme provides long, consistent tracks. Now this gives us a solid base for BA initialization. Thus, in this section we evaluate the proposed method with its best performing settings. We compare the method implemented with and without local BA to the baseline method [Richa et al., 2014]. We use LCE metric introduced in Chapter 4. This shows how the composition of estimated transformations affects the global registration and accumulated drift. The idea is to initialize a uniform grid of points $g_1, ..., g_{n_l}$ at the first frame of the sequence and use the set of pairwise estimated transformations applied sequentially to transfer the grid through the sequence.



(a) (b) (c)

(d) (e) (f)

Figure 5.6: Examples of improved areas of the mosaics given in Figure 5.1 with corrected drift using proposed approach. First column - originals, second column - corrected versions.

The metric computes the discrepancy between the initial and resulting sets of points as:

$$\xi_{LCE} = \sqrt{\frac{1}{l}\sum_{i=1}^{n_l} \| g_i - \zeta_i \|_2^2} \tag{5.6}$$

where $\zeta_i = w(...(w(g_i, \theta_{1,2}))..., \theta_{f,1})$. The comparison of our method and [Richa et al., 2014] are shown in Table 5.4. One can see, that our method outperforms the baseline method. A significant improvement can be observed on the version of the proposed method where the local BA step was used.

	dataset#1	**dataset#2**	**dataset#3**	**dataset#4**
proposed (1)	30.43	21.75	48.02	49.12
proposed (2)	11.36	5.48	32.16	38.56
baseline [Richa et al., 2014]	34.18	28.64	48.15	50.72

Table 5.4: LCE computed across datasets. The proposed (1) is our method with UGrid used for tracks initialization and SHD based local correction step. The proposed (2) is the proposed (1) + local BA.

Illustration of the improvement achieved on the sample mosaic shown in Figure 5.1 is given in Figure 5.6. Cropped regions of interest before the application of our drift reduction mosaicing method are given in the first row and the corrected versions are given in the second row. One can see the vessel misalignment initially present in Figure 5.6a was corrected and the vessel remains continuous, as shown in Figure 5.6d. The blurred vessel in Figure 5.6b and duplicated one from Figure 5.6c were also corrected and visual quality has been improved as it is shown in Figures 5.6e and 5.6f respectively.

5.4 Conclusion

In this chapter we presented a method for drift reduction in mosaicing slit-lamp retinal video sequences. We validated it using a simple global motion model that can efficiently produce long-term tracks with a better precision for the long video sequences. We also demonstrated that using a grid of points distributed uniformly over the visible part of the retina generally provides a better initialization for tracking. We have proposed a new local refinement procedure which potentially may be successfully applied within the scope of other applications such as object tracking in the non-medical domain. This, however, was not tested.

Chapter 6

Handling Reflection Artifacts

In this chapter we propose an effective technique to detect and correct light reflections of different degrees in SLIM. This serves our second thesis objective - enhancement of the global photometric quality of the resulting mosaic by minimizing the illumination artifacts. We start by explaining our motivation in §6.1. The description of our two stage method is given in §6.2. We explain how the specular-free image concept can be used to obtain glare-free image and use it coupled with a contextually driven probability map to segment the visible part of the retina in every frame before image mosaicing. We also demonstrate the steps required to perform a new label-specific blending that takes into account the types of specular highlights. We also introduce a new quantitative measure for global photometric quality. Evaluation results on a set of video sequences obtained from slit-lamp examination sessions of 11 different patients presenting healthy and unhealthy retinas are given in §6.3 with corresponding discussion and we summarize the chapter in §6.4.

Contents

6.1 Motivation **70**

6.2 Methodology **71**

6.2.1 Single-image glare removal and retina segmentation 71

6.2.2 Multi-image lens flare correction: content-aware blending 72

6.3 Experimental results and discussion **74**

6.3.1 Dataset and evaluation strategy 74

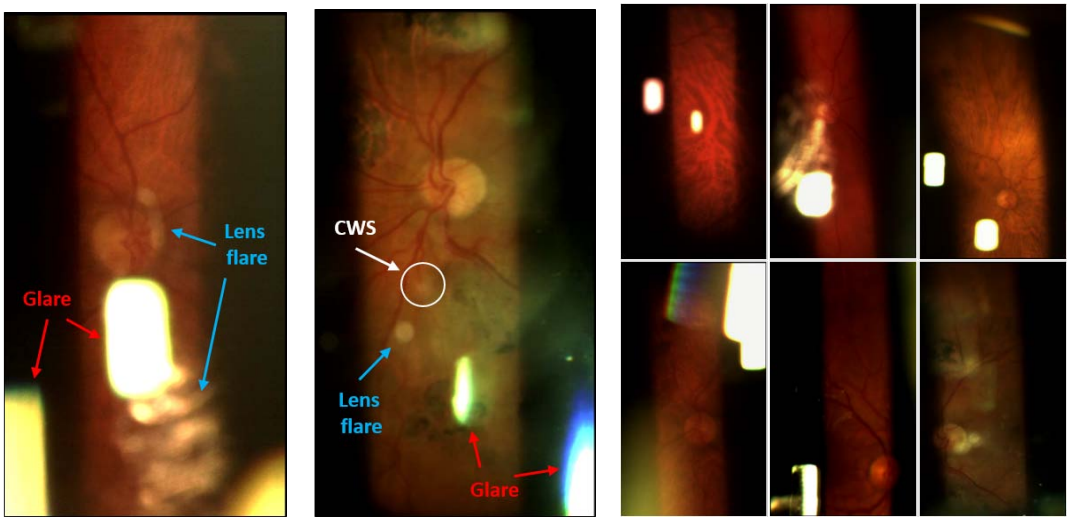
6.3.2 Single-image glare removal and retina segmentation 74

6.3.3 Multi-image highlight correction: content-aware blending 77

6.4 Conclusion **77**

6.1 Motivation

Obtaining a geometrically and photometrically accurate retinal mosaic in SLIM is a difficult task due to the numerous challenging conditions. When performing retinal examination with a slit-lamp the imaging set-up is arranged so that the axis of the observation component is nearly coaxial with the axis of the illumination component. Both are fixed on the moving base, controlled by the ophthalmologist. The light beam is focused on the retina using a hand-held direct contact lens of strong convergence. This essential requirement, unfortunately, introduces bothersome illumination artifacts that populate the image as illustrated in Figure 6.1. The reflections of the light from the slit-lamp on the cornea (the transparent layer forming the front of the eye) and the contact lens create specular highlights that are difficult to separate from the retina. An example of this is shown in Figure 6.1a. Worsened by changing exposure which is adopted to the patient's comfort, they obscure and degrade retinal details. Moreover, as they appear brighter than the dominant color of the retina they may be wrongly recognized as 'cotton wool spots' - abnormal findings on the retina which appear as small, yellow-white (or grayish-white), slightly elevated cloud-like lesions. An example of this is shown in Figure 6.1b. These are typical for diabetic retinopathy, hence complicating diagnosis.



(a) Healthy retina (b) Diabetic retina (c) Inter-patient appearance variation

Figure 6.1: Typical slit-lamp images demonstrating the appearance variation of the light reflection of different origins. CWS - cotton wool spot.

Reflection control during slit-lamp assisted examination relies primarily on intentional 'light fallof' (*i.e.*reduction of the light intensity to provide a better comfort to the patient, that results in darkening of image corners) and anti-reflection coatings of the contact lens. Despite this provision, the unwanted reflections still occur. Glare eliminates all information in the affected pixels and the other types of reflections can introduce artifacts in feature extraction algorithms, which are critical in our application. Existing solutions in SLIM, as discussed in the previous work §3.4, manage to deal with strong glares which corrupt the retinal content entirely while leaving aside the correction of semi-transparent specular highlights and lens flare. This introduces ghosting and information loss. Moreover, popular generic methods share two common problems: *1)* they generally result in noticeable artifacts when applied directly in SLIM and *2)* the motion cues, utilized in multi-view approaches, are useful but cannot be used as hard constraints because apparent motion of specular highlights can be noticed, but, unlike in previous work, more than two consecutive observations are required. Hence, a new methodology to overcome the aforementioned issues and obtain visually consistent mosaics in SLIM is on demand.

6.2 Methodology

We outline the proposed method in Figure 6.2. This consists of two main stages: *(i)* single-image glare removal and retina segmentation and *(ii)* multi-image lens flare correction by content-aware blending. We rely on the SF image concept introduced in [Tan and Ikeuchi, 2005] to obtain Glare-Free (GF) image and use it coupled with contextually driven probability maps to segment the visible part of the retina at every frame before image mosaicing. For the sake of clarity we show the mosaicing block in a compact form and refer to the mosaicing method with drift reduction described in Chapter 5. We then proceed directly to the image blending where a subset or all frames have been transformed and spatially aligned. We detect the lens flare areas on a set of overlapping images and label each pixel as 'flare' or 'non flare' using a probability maps. Finally, we invoke an adequate blending method.

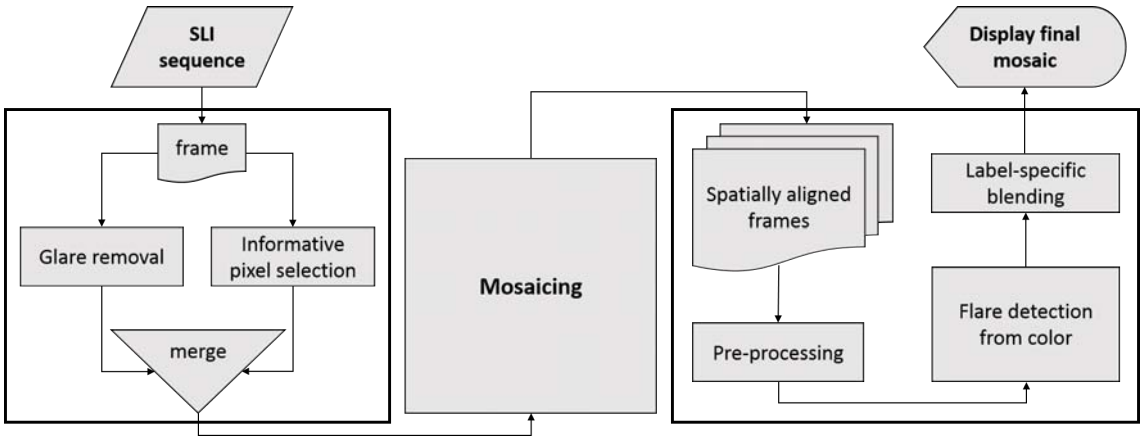


Figure 6.2: Flowchart of the method to correct light-related imaging artifacts in SLIM.

6.2.1 Single-image glare removal and retina segmentation

Segmentation of the informative retinal content from slit-lamp images is a challenging task. The recent work of [Zanet et al., 2016] was not found to be suitable for effective retina segmentation in our datasets. The concept of specular-free image used in related works is just an approximation of the ground truth. Nonetheless, it has been demonstrated to be effective for single-image glare removal. Incorporating contextual information is considered as one of the most effective approaches in medical applications [Collins et al., 2014]. Retinal images obtained with a slit-lamp have a narrow FOV localized in the center of the image resulting in big part of the image containing dark pixels. This property can be used to obtain the Region of Interest (ROI) to reduce the processing load. Our approach can be summarized in three steps. The schematic illustration is shown in Figure 6.3 and the detailed description of every step is given as follows.

Step 1: pre-processing. First the image is converted to the LMS (Long, Medium, and Short light wavelengths) color space. This is commonly used color space to estimate the appearance of a pixel under a different illumination. Based on the observation that the maximum fraction of the unsaturated pixels in local patches changes smoothly we proceed with low-pass filtering similar to [Yang et al., 2010] and obtain I_{LP}. We then compute $C_{min} = \frac{min(I_{LP})}{mean(I_{LP})}$ - the maximum chromaticity image as a pixelwise division of the minimum value over three components of I_{LP} and the mean value respectively. This computation results in a binary image, where the most glare pixels have intensity equal to 1.

Step 2: informative pixel selection. Given the priors on the location of the slit in the image we filter out highly improbable locations of the informative retinal content. For each pixel in the image

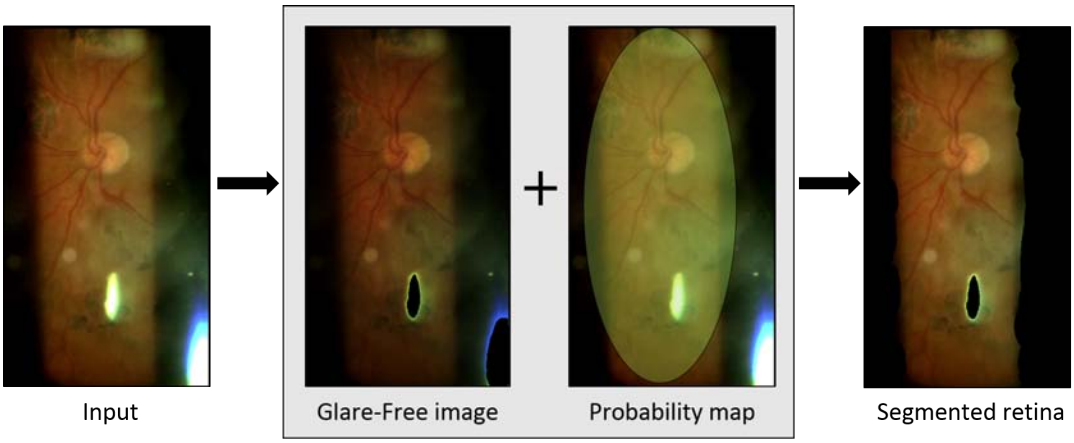


Figure 6.3: Schematic illustration of the single-image glare removal and retina segmentation in SLIM.

we compute a conditional probability of the retinal content occurring at this pixel given the center of the image $\mathbf{c}_I$. We model this contextual constraint with a Gaussian Mixture Model (GMM):

$$P(\mathbf{p}|\mathbf{c}_I) = \sum_{i=1}^{K} w_i G(\mathbf{p} - \mathbf{c}_I; \mu_i, \Sigma_i) \tag{6.1}$$

where K is the number of GMM components and $\{w_i, \mu_i, \Sigma_i\}$, $i = 1, ..., K$ are the GMM's parameters estimated with Expectation Maximization (EM). The model is learned offline on a set of annotated frames from different video sequences. Here, K was empirically tuned to represent two Gaussian components. We apply the model on a test frame and obtain a probability map.

Step 3: combination. Here we incorporate the positional prior learned in the previous step to filter out uninformative areas of the image and obtain the final segmentation of retinal content. Thus, we keep $\mathbf{p}$ as a retinal content if $P(\mathbf{p}|\mathbf{c}_I) \geq t$, where t is a probability threshold which we empirically set to 0.6. We perform a logical XOR operation with the GF image mask from *Step 1* within the estimated region. This allows us to keep only those pixels, where the GF mask or the estimated region, but not both, contain a nonzero element at the same location.

6.2.2 Multi-image lens flare correction: content-aware blending

Localized flare patches in areas of uniform color and brightness in non-medical images can be easily corrected by copying parts of neighboring areas over the affected area. The situation is much more complicated when the flare affects areas with lots of detail and tonal variations as retinal content. Correction is generally not possible without knowing beforehand what the affected areas should look like in the absence of flare. This requires a sophisticated per-pixel analysis in different views. Given a set of spatially aligned images we want to detect which pixels are likely to be pixels affected by lens flare. Once the lens flare regions are revealed, their visibility may be corrected by performing an adequate color mapping. The procedure can be summarized in three steps. The schematic illustration is shown in Figure 6.4 and the detailed description of every step is given as follows.

Step 1: pre-processing. Because reflection caused by lens flare has a complicated nature it is necessary to address the problem within an appropriate color space representation. Thus, for a given pixel on the mosaic $M(\mathbf{q})$, a set of overlapping frames are first transformed to the L*a*b color space. Following the same reasoning as described in the *step 1* in the previous section, we apply an image guided filter to the L component. Because the L component represents scene

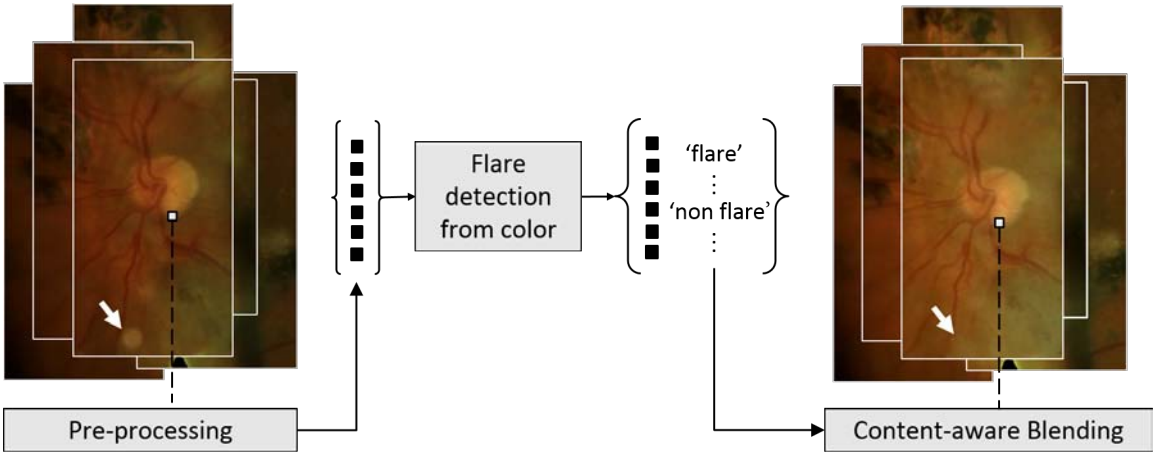


Figure 6.4: Schematic illustration of the multi-image lens flare correction in SLIM.

luminance and low-pass filtering adjusts the local intensity to its neighbors it is more likely to obtain well preserved boundaries of areas affected by lens flare.

Step 2: flare detection from color. Because regions affected by lens flare have specific colors, which are different from the rest of the retina, it motivates the use of color GMMs. We learn a simple GMM similarly to [Chhatkuli et al., 2014] offline on a set of manually annotated images where the pre-processing from the previous step was applied:

$$P(\mathbf{l}|\lambda) = \sum_{i=1}^{K} w_i G(\mathbf{l}; \mu_i, \Sigma_i) \tag{6.2}$$

with $K = 3$ Gaussian components. Here $\mathbf{l}$ is the image pixel and $\lambda = \{w_i, \mu_i, \Sigma_i\}$, $i = 1, \ldots, K$ are the GMM's parameters estimated with EM. We obtain a probability map for every L component in the observation set of frames on the mosaic using the trained GMMs. This indicates the probability that a given pixel in the observation belongs to the flared region. We use a Graph Cut algorithm [Boykov and Jolly, 2000] to mark the pixel as 'flare' or 'non flare'. As this is posed as a binary labeling problem, the Pott's Energy function is sufficient:

$$E(I) = \sum_{\mathbf{p} \in S} |I_{\mathbf{p}} - I'_{\mathbf{p}}| + \sum_{\mathbf{p},\mathbf{q} \in N} P(\mathbf{p},\mathbf{q}) T(I_{\mathbf{p}} \neq I_{\mathbf{q}}) \tag{6.3}$$

where $I = \{I_{\mathbf{p}} | \mathbf{p} \in S\}$ are the unknown labels over the set of pixels S and $I' = \{I'_{\mathbf{p}} | \mathbf{p} \in S\}$ are the observed labels. The Pott's interaction is specified by $P(\mathbf{p},\mathbf{q})$, which are the penalties for label discontinuities between adjacent pixels. The function T is an indicator function. This is optimally solved by a single execution of max-flow.

Step 3: blending. We count the number of pixels belonging to each label and identify the majority. We take the average luminance L of the majority as a L_t - *top luminance* and the average of the rest of the pixels as a L_b - *bottom luminance*. We then invoke an appropriate mapping function. This is inspired by [Reinhard et al., 2002]. Thus, if the majority is 'flare' pixels we apply 'color burning' - divide the inverted L_b by the L_t, and then invert the result as $C_{burn} = 1 - (1 - L_b)/L_t$. This darkens the L_t increasing the contrast. In the opposite case we apply 'color dodging' - divide the L_b by the inverted L_t such as $C_{dodge} = L_b/(1 - L_t)$. This lightens the L_b depending on the value of the L_t.

6.3 Experimental results and discussion

6.3.1 Dataset and evaluation strategy

The datasets used for evaluation were obtained from slit-lamp examination sessions performed on 11 different patients at University Hospital of Saint-Étienne, France, presenting healthy and unhealthy retinas. The proposed glare removal and retina segmentation were evaluated on a set of 270 manually annotated image frames sampled from the set of videos. This was to ensure the coverage of patient-specific and lens-specific specular highlight variation. The images were annotated with binary masks to separately assess the performance of glare removal and retina segmentation. The proposed blending technique for lens flare correction was rated on a set of geometrically aligned video frames obtained by the mosaicing method with drift reduction described in Chapter 5. For simplicity we will further refer to it as SLIM-DF.

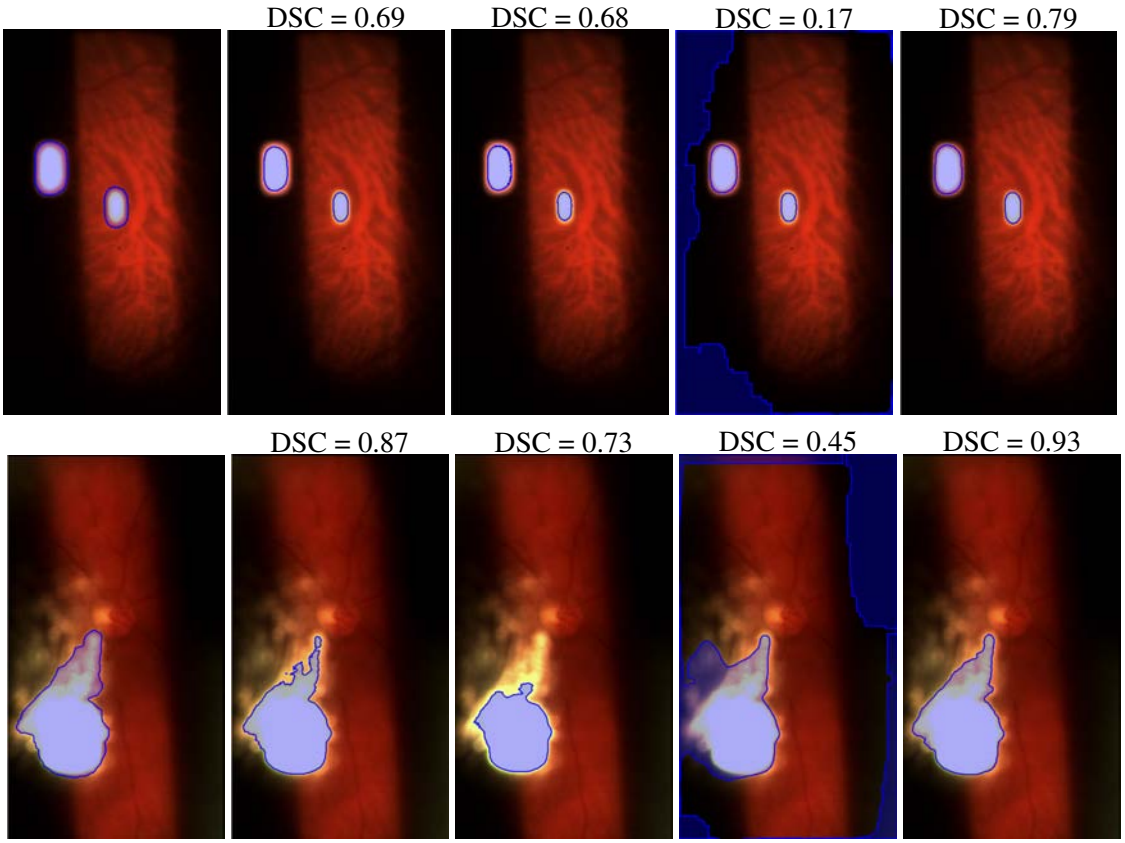
6.3.2 Single-image glare removal and retina segmentation

We start with the comparison of our glare removal technique with the existing methods proposed in [Tan and Ikeuchi, 2005, Shen et al., 2008, Yang et al., 2010]. We manually annotated selected datasets by drawing the contour around regions obscured by highly saturated pixels. In simple cases, where the patient appeared to be less photosensitive and the image acquisition was not polluted by mixture of different degrees of reflections the glared region boundaries were easy to locate. Because most of the time it is difficult to observe a clear boundary between a glared region and the surrounding distorted areas, we opted for a middleground. The results for such two cases are shown in Figure 6.5. We computed the DSC to assess the similarity with the annotated regions. The higher the value the more similar the algorithm's output to the reference mask. For the simple case shown in the first raw, all the methods perform well while in the difficult case, shown in the second row, only the proposed method provides acceptable results.

We then combine the GF image with the spatial probability map to obtain the visible retinal content. The experimental results of our method compared to the simple thresholding with morphological refinment that we used in SLIM-DF §5.2.1 and the ML-based approach proposed in [Zanet et al., 2016], are illustrated in Figure 6.6. Here we also compute statistical measures for every output and average it over results on 270 annotated samples as shown in Table 6.1. One can see that our method provides higher values indicating better performance.

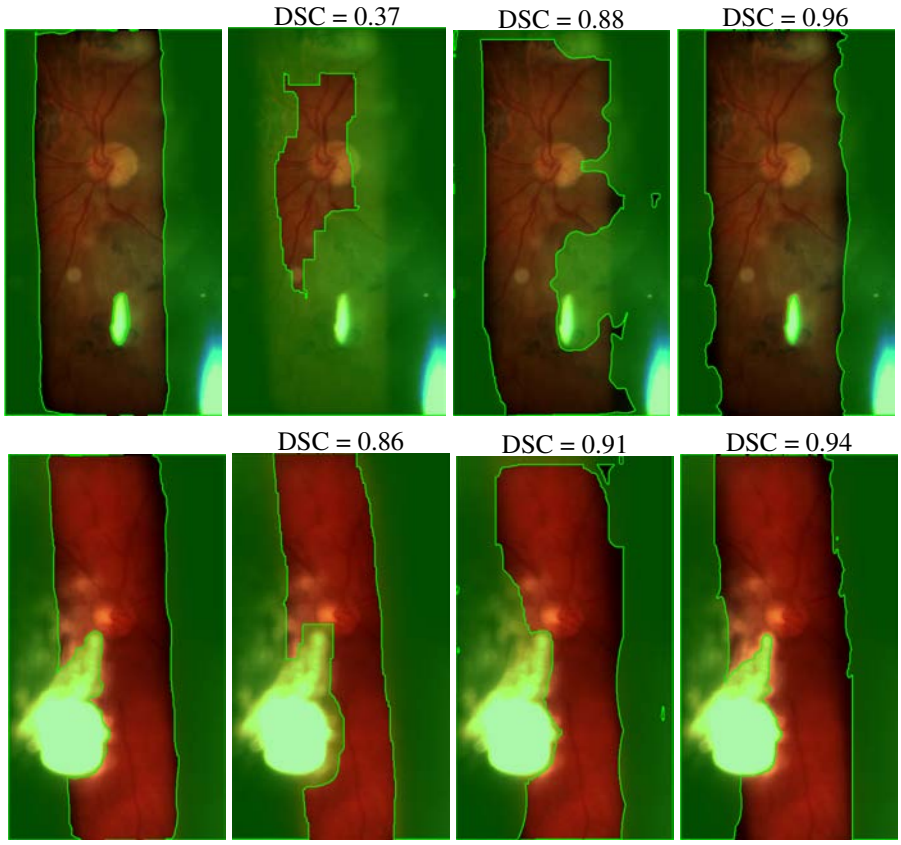
	Precision	**Accuracy**	**Specificity**	**Sensitivity**
Thresholding as in SLIM-DF	0.30	0.70	0.58	0.86
Method [Zanet et al., 2016]	0.90	0.92	0.94	0.89
Proposed	0.92	0.95	0.97	0.90

Table 6.1: Retinal content segmentation performance.



(a) GT image (b) Method 1 (c) Method 2 (d) Method 3 (e) Proposed

Figure 6.5: Comparative results of glare removal. GT - Ground Truth. Example of a simple case is shown in the first row and the second row illustrates more complicated condition. Method 1 - [Tan and Ikeuchi, 2005], Method 2 - [Shen et al., 2008], Method 3 - [Yang et al., 2010].



(a) GT (b) Method 1 (c) Method 2 (d) Proposed

Figure 6.6: Comparative results of retinal content segmentation. GT - Ground Truth. Method 1 - thresholding as in SLIM-DF and Method 2 - [Zanet et al., 2016].

6.3.3 Multi-image highlight correction: content-aware blending

The most traditional way to evaluate the photometric quality of slit-lamp image mosaics is still based on the visual assessment of ophthalmologists. Even though the experts' opinion is a good reference it is a subjective evaluation which may differ between experts and may prevent the mosaic from being used. Here we propose a new quantitative evaluation of the global photometric quality. We propose to use a Blending Consistency Measure (BCM). It assesses the quality of the blending by computing the standard deviation of a pixel's intensity in the transformed image $I(\mathbf{q})$ from a set of corresponding locations in the mosaic $M_i, i = 1, 2, ..., n$ as:

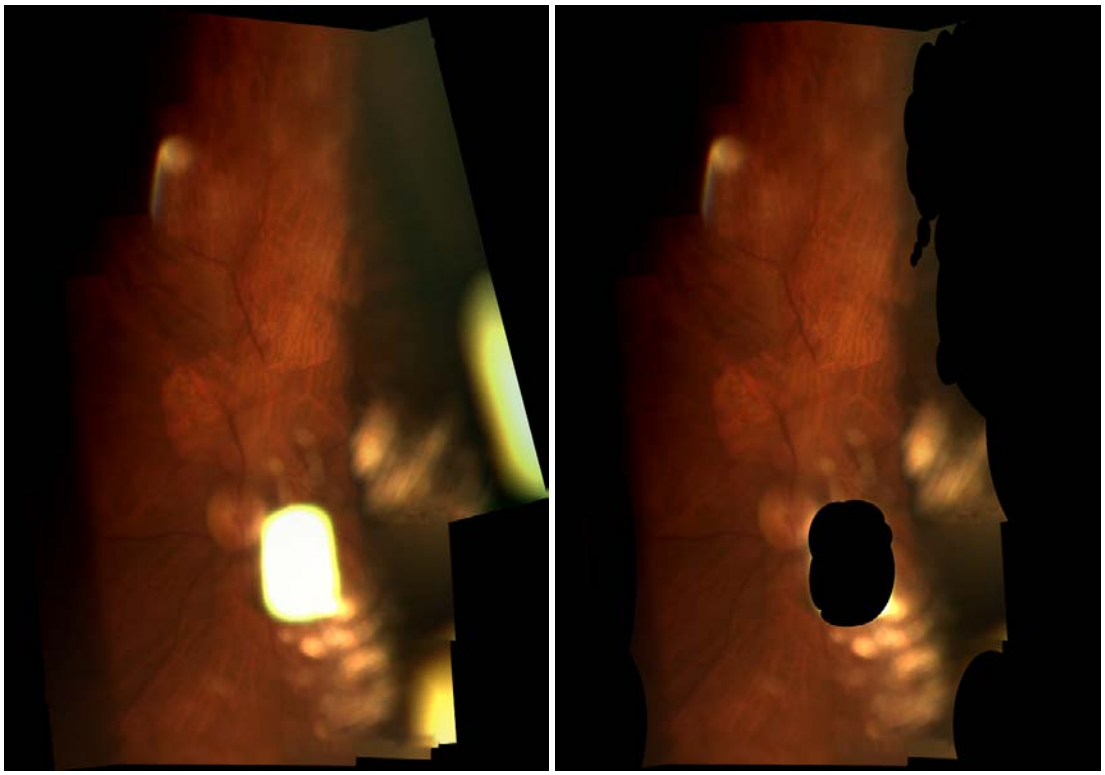
$$BCM = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \mid I(\mathbf{q}) - \mu \mid^2} \quad \text{, where} \quad \mu = \frac{1}{n}\sum_{i=1}^{n} M_i \tag{6.4}$$

The results shown in Figure 6.7 demonstrate one of the mosaics for visual assessment. We take the mosaicing result obtained by the modified version of the SLIM-DF, where we removed the illumination correction. We then compute BCM for this uncorrected mosaic and the results obtained with the inclusion of the correction techniques from existing work in SLIM [Zanet et al., 2016]and the proposed method. The computed metric spans the range [0;255]. We show the computed results represented as a percentage value. The smaller the value, the better the blending consistency. The mosaic shown in Figure 6.7c consists of 530 frames while mosaics shown in Figures 6.7a, 6.7b and 6.7d are made of 212 frames as they are based on SLIM-DF which uses key-frames. As can be noticed, the result in Figure 6.7c appears darker compare to the others. This is due to the blending method used in [Zanet et al., 2016], where the intensity fades toward the border of the segmentation mask which we have re-implemented strictly following the provided formulas.

One can see that the proposed method significantly improves the global photometric quality of the mosaic in the major areas and outperforms existing works. This is true for the majority of the cases in our dataset. However, it does not work well in the lower right corner of the illustrated example shown in Figure 6.7d. Our glare removal part was specifically designed to work for a middleground and keep as much more valuable information as possible. Thus, it is not always able to erase all the glare-like artifacts but it always keeps the 'uncorrupted' retina. [Zanet et al., 2016], in the other hand, does not include the mentioned region as shown in Figure 6.7c and, according to the experiments on other sequences, it cuts out a big part of the retinal content which is not corrupted by artifacts and can be useful for diagnostic purposes. This complication may be due to various reasons: different contact lenses were used in the procedure, the manual navigation by an ophthalmologist is not always precise, and the industrial prototype we use is constantly under development and it is not perfect. The improvement we expect to achieve in future work will mainly come from the improvement of the prototype itself. More examples of mosaics are shown in Figure 6.8.

6.4 Conclusion

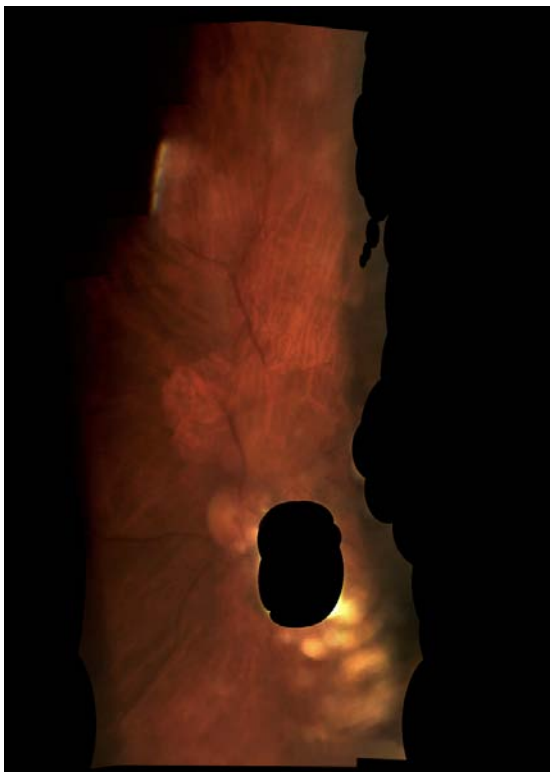
In this chapter we showed how to segment the informative part of the retinal content and correct specular highlights of different degrees in SLIM. To this end we studied several specular highlight removal and correction approaches proposed in the medical and non-medical domains and designed our own solution specifically adapted to our task. Firstly, we improve on the previous works by proposing a fast single-image technique to remove glares and segment the visible retina using the concept of specular-free image and contextual information. Secondly, we incorporate the notion of the type of specular highlight and motion cue for intelligent image blending. Our experimental results showed that the proposed methodology exhibits a good efficiency, significantly outperforming related works in SLIM.



(a) No correction, $BCM = 63\%$

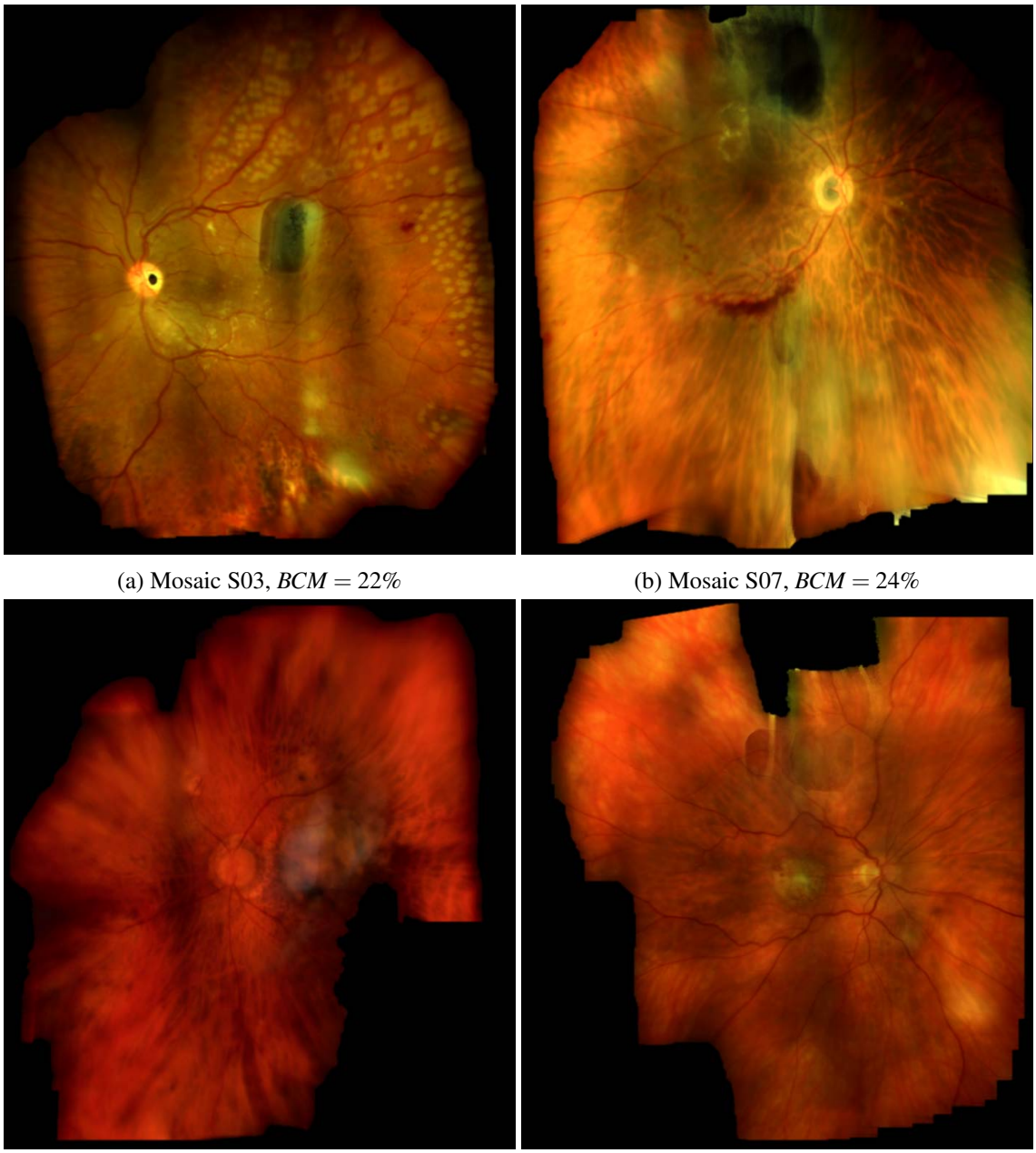
(b) Method 1, $BCM = 51\%$

(c) Method 2, $BCM = 38\%$



(d) Proposed $BCM = 23\%$

Figure 6.7: The comparative results for one of the mosaics. Method 1 - SLIM-DF and Method 2 - [Zanet et al., 2016].



(a) Mosaic S03, $BCM = 22\%$

(b) Mosaic S07, $BCM = 24\%$

(c) Mosaic S11, $BCM = 15\%$

(d) Mosaic S08, $BCM = 7\%$

Figure 6.8: Example of mosaics build with SLIM-DF and corrected with the proposed technique.

Chapter 7

Angio2SLIM: Automatic Multimodal Registration

In this chapter we present our last contribution - a novel fully automatic multi-modal registration method called Angio2SLIM, which automates the process of registering Fluorescein Angiography images and SLIM. We start by explaining our motivation in §7.1. The majority of existing methods require a detection of common feature points in both image modalities. This is a very difficult task for SLIM and FA. In addition, they do not account for the accurate registration in the macula area - the priority landmark. Moreover, none has attempted to achieve a fully automatic solution for SLIM and FA before. We describe the proposed method in detail in §7.2. Our solution is built upon an unsupervised iterative stochastic optimization where the point correspondences are established collaboratively. A data-driven measure of correspondence quality is used which combines texture, spatial information, rotation and illumination invariance. The final registration is achieved by fitting the normalized quadratic model. We present experimental validation, both qualitative and quantitative, on multiple patient datasets in §7.3. We show that our method provides more accurate registration compared to the semi-automatic baseline registration. We summarize the chapter in §7.4 and give an overview of the future work.

Contents

7.1 Motivation **82**

7.2 Angio2SLIM **83**

7.2.1 Retinal features detection in a reference FA image 83

7.2.2 Automatic matching using SOM and LBP 85

7.2.3 Non-rigid image registration with the normalized quadratic model 89

7.3 Experimental results and discussion **90**

7.3.1 Datasets and ground truth 90

7.3.2 Inclusion of the priority landmark detection 92

7.3.3 Automatic point matching 93

7.3.4 Multi-modal registration: comparative results 94

7.4 Conclusion **98**

7.1 Motivation

Automatically registering SLIM with diagnostic images of other modalities such as Fluorescein Angiography allows the ophthalmologist to prepare the treatment plan by precisely indicating the areas for laser application and zones where no intervention is required. Such a multi-modal registration is not an easy task, however, due to the large variability, both in geometry and texture. Figure 7.1 shows several sample images from both modalities for the reference. Moreover, it implies that a special care should be taken over the priority landmark - the macula region, located in the central area of the retina. Because it is responsible for the high-resolution central color vision, accidental damage caused by inaccurate laser burns may cause total blindness. In addition, the detection of anatomical landmarks in both images is not trivial because the same landmark detector does not work well on both modalities, requiring one to employ two different methods. This reduces the generalization potential to other retinal image modalities. Finally, a semi-automatic method, currently implemented in TrackScan, requires a human operator to select point correspondences between images. It then initiates the automatic search of an optimal set for the rigid model transformation parameters. Assuming a rigid body transformation in this case does not reflect the complex deformation field that exists between the images. Even if the patient's retina is less likely to change significantly between image acquisitions distant in time, a more complex transformation model shall be assumed to account for the SLIM modality as it is not a 'one-shot' image but the result of image composition achieved by affine deformations. It is also beneficial to automate the registration process so that no human operator is involved.

These issues can be solved by the use of SOM to automatically establish point correspondences coupled with landmark detection in the reference image only. As we saw in the discussion of the previous work in §3.3.2, the application of SOM to the problem of multi-modal image registration has already been studied [Matsopoulos et al., 2004, Markaki et al., 2009]. The gradient difference metric used to establish point correspondences in this method, however, does not perform well on our dataset. Thus, it would be interesting to see if SOM with a more suitable similarity measure can handle such challenging multi-modal case as registration of SLIM and FA. Besides, the detection of vessel bifurcations in FA is relatively simple, while it remains much more difficult for SLIM due to the presence of motion blur and specular noise induced by the acquisition system. Thus, FA is a natural choice for the reference image. There exist different contextual constraints between the

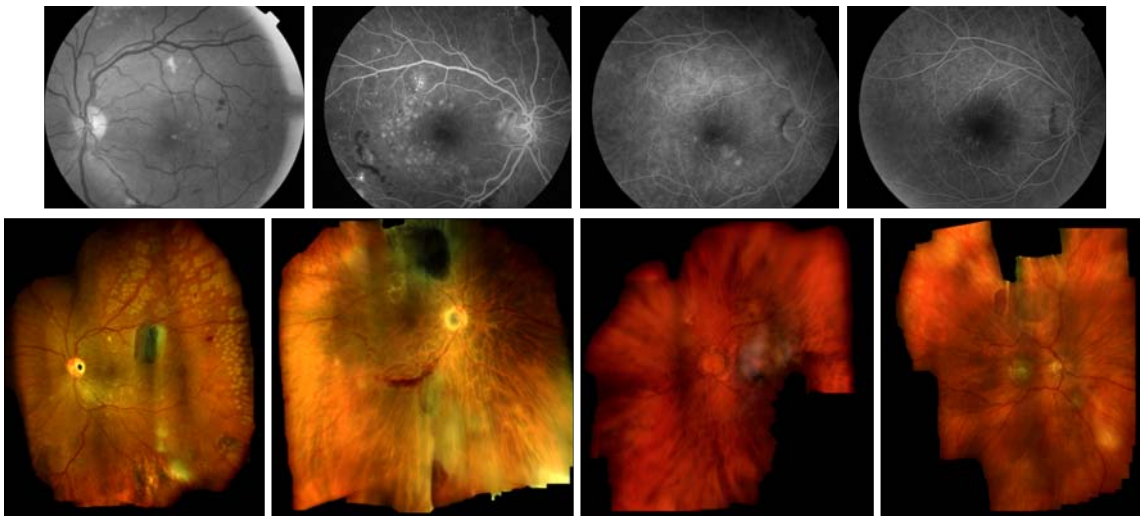


Figure 7.1: Illustration of the inter- and intra-modality geometric and photometric variability between FA (first row) and SLIM (second raw).

location of the optic disc and the macula. Additionally, in SLIM the retina is mapped in such a way that the Optic Disc (OD) is located in the central area of the image. These positional priors allow us to perform macula localization without introducing more complexity and do pre-alignment to provide a decent initialization for registration. Finally, an affine transformation is preferable for matching because it only needs to be valid locally. A global registration can then follow using the quadratic model [Can et al., 2002].

7.2 Angio2SLIM

A schematic overview of the proposed framework is illustrated in Figure 7.2. It consists of three stages: *(i)* retinal features (vessel bifurcations) detection on the reference FA image complemented by macula localization and extraction of SURF features in the macula area, *(ii)* automatic identification of feature correspondences using SOMs and *(iii)* non-rigid image registration performed using the quadratic model and the feature correspondences from the previous stage. We restrict the scope of the problem to images where all types of retinal anatomical landmarks (*i.e* optic disc and macula) are fully visible. A detailed description is provided in the following sections.

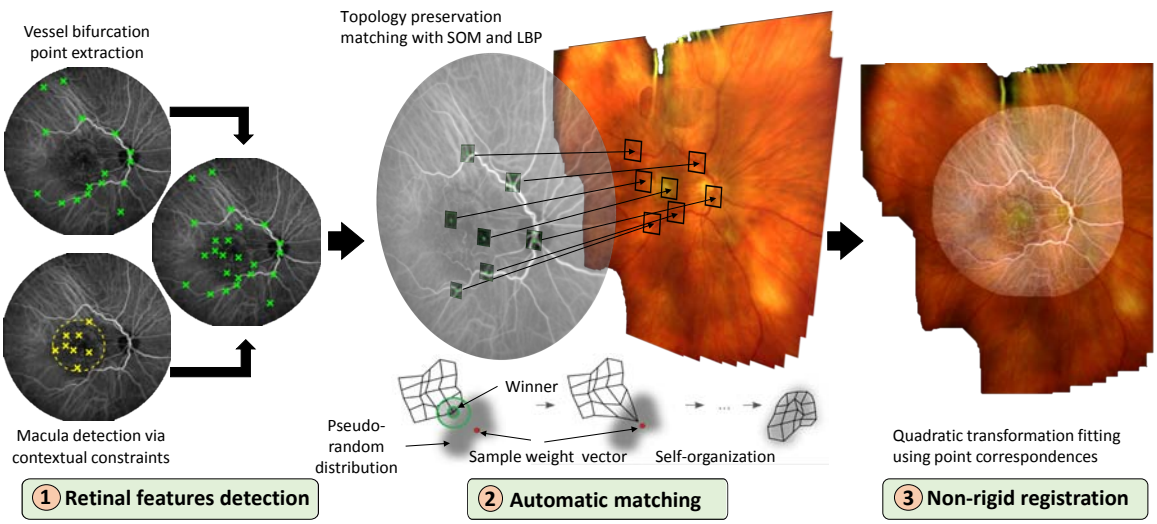


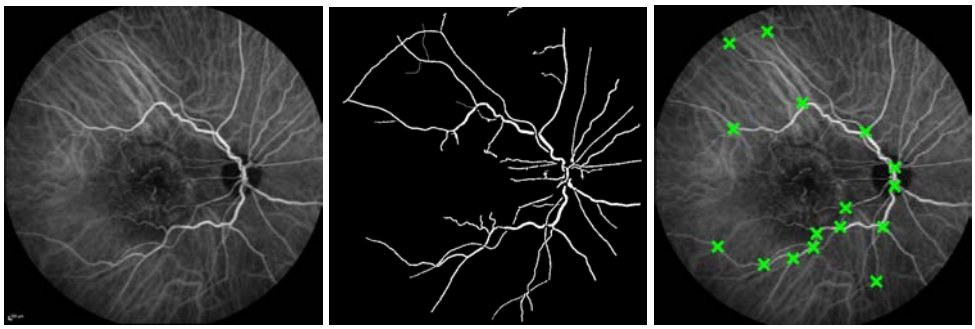
Figure 7.2: Method overview.

7.2.1 Retinal features detection in a reference FA image

A good image registration process requires that a sufficient number of uniformly distributed corresponding feature points are present in both images. For these reasons, the blood vessel bifurcations form a natural choice. Nevertheless, the detection of these features using the available solutions such as [Can et al., 2002] or [Choe and Cohen, 2005], for example, pose significant difficulties when applied to SLIM. However, recent solutions for retinal vessel segmentation based on the concept of deep learning, such as [Maninis et al., 2016], provide very competitive result on fundus photographs, but their direct application to SLIM fails and fine-tuning of the available pre-trained model is out of the scope of this work. Thus, we obtain bifurcation points from the FA image only. Relying solely on bifurcation points, however, does not ensure an accurate registration result in the macula area. Thus, we propose to complement bifurcation point detection with macula localization based on contextual

constraints, and to extract additional interest points from that region. The following paragraphs give a detailed description to our approach.

Vessel bifurcations detection Many of the techniques proposed to extract vessel bifurcation points from FA-like images do not provide publicly available implementations. The design of yet another method for this purpose is not our goal. Thus, we proceed with a simple approach that is capable of providing us what we need, as shown in Figure 7.3. We first compute a complement of the FA image A_c by subtracting each pixel value from the maximum image intensity $A_c(x,y) = max(A) - A(x,y)$. In the output image, dark areas become lighter and light areas become darker. We proceed with background subtraction, where a mean filter is applied to the contrast enhanced image. The contrast enhancement function $\delta(A_c)$ here is the Contrast Limited Adaptive Histogram Equalization (CLAHE) [Zuiderveld, 1994] method. Intensity thresholding is then applied where the Iterative Self-Organizing Data Analysis Technique (ISODATA) [Ridler and Calvard, 1978] method is used to determine the global image threshold. The vessel tree is already segmented at this point but surrounded by some spurious pixels which we remove with morphological opening. We detect bifurcation points as the branch points of the skeletonized vessel tree from the segmented image. We use the implementation of Contour-Pruned Skeletonization from [Howe, 2015] and a look-up table to locate branch points on the skeleton.



(a) Input image (b) Segmentation result (c) Bifurcation points

Figure 7.3: The simple approach used to obtain vessel bifurcation points in the FA image.

Optic disc localization The macula can be effectively localized based on its distance and position with respect to the OD's center. Therefore we first perform the OD localization and infer macular region using OD location prior. Despite the intensity variations, orientation of the image and imaging artifacts, the OD's shape is mostly circular and it has a dense amount of retinal vessels converging to its center. The macula area, in contrast, does not contain any distinctive features. Because the localization of the OD is much easier to find out compared to the macula area, and the OD is always fully visible in the images from our dataset, we opted for a template matching technique, as illustrated in Figure 7.4. This consists of two stages that can be summarized as follows:

Training we manually select a square region of size $R = 50$ pixels around the OD for every FA image in our dataset. We apply a rotation transform to randomly chosen images covering 50% of the data to ensure rotational invariance as some of the FA images may not be in a strictly up-right orientation. We then compute the average image I_O^T to be used as a template.

Testing we apply an exhaustive search and compute the NCC coefficient between I_O^T and all the pixels in the image. We accept the maximum NCC response as the nominal center C_O of the OD . This gives us a contextual prior for the localization of the macula area.

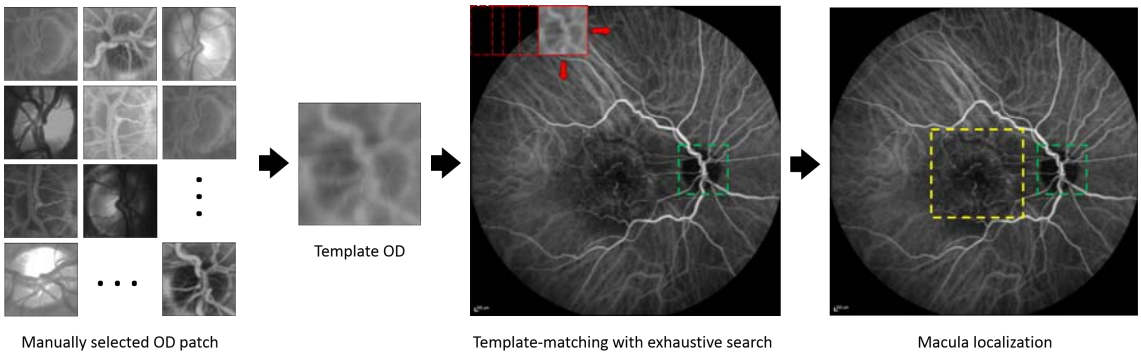
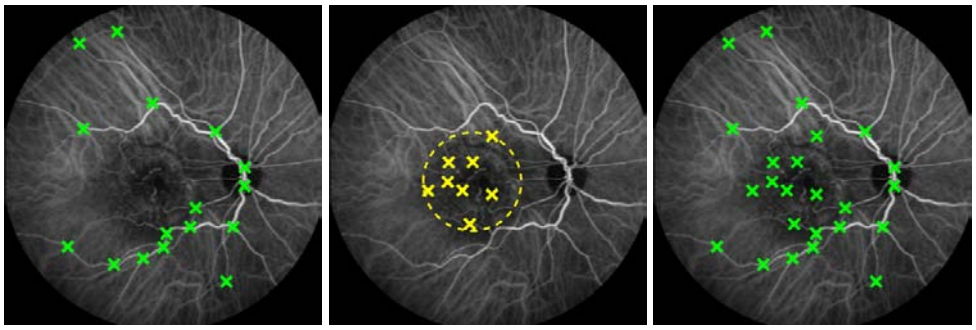


Figure 7.4: Template matching scheme for the OD and macula localization.

Macula localization and SURF features extraction Because the location of the macula varies among patients and we do not aim to precisely obtain its contours, a rectangular area is sufficient for our task. In a standard FA image the macula is situated about two and a half disc diameters to the left or right of the disc. Thus, we take the value R used for disc detection as a rough diameter estimate and use it to localize the center of the macula region as $C_M = C_O \pm 2.5R$. To identify the right side for the macular region placement, we check if C_O falls to the left or right side from the image center. Finally, we extract SURF features from the local neighborhood of the macula center and append them to the bifurcation points to form the final set of retinal features $\mathbf{p}_i \in \mathbb{R}^2, i = 1, ..., N$ as shown in Figure 7.5.



(a) Bifurcation points (b) SURF macula points (c) Joint feature set

Figure 7.5: Example of a result of the retinal features detection in the FA reference image.

7.2.2 Automatic matching using SOM and LBP

Automatically establishing correspondences between images is a central problem in multi-modal registration. To this end, a self-organizing process is a natural and very promising setting. Our algorithm is an adaptation of the method proposed by [Matsopoulos et al., 2004] and extended in [Markaki et al., 2009] where the SOM is used as a basis. It has proved to be strongly resilient to outliers, it is not critically influenced by the control parameter selection, and less exposed to the effects of multi-modality and local optima [Matsopoulos et al., 2004, Markaki et al., 2009].

Self-organizing Maps SOM is a neural network algorithm, which uses a competitive learning technique to train in an unsupervised manner. SOMs differ from other artificial neural networks as they apply competitive learning as opposed to error-correction learning, and in the sense that they use a neighborhood function to preserve the topological properties of the input space. [Kohonen, 1998] first established the relevant theory and explored possible applications. The Kohonen's model comprises a

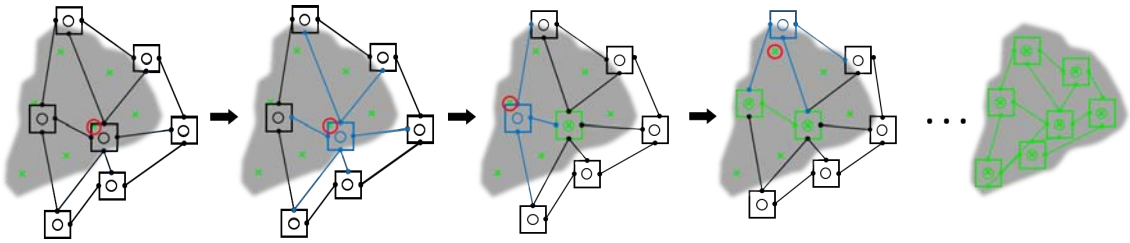


Figure 7.6: An illustration of the training of a SOM.

layer of neurons, ordered usually in a one-dimensional or 2D grid. The goal of learning in the SOM is to force different parts of the network to respond similarly to certain input patterns. The weights of the neurons are initialized either to small random values or sampled evenly from a predefined space. With the latter alternative, learning is much faster because the initial weights already give a good approximation of SOM weights. The network receives a large number of example vectors during training that represent, as close as possible, the kinds of vectors expected during mapping. The examples are usually given iteratively. The training utilizes competitive learning. When a randomly chosen training example is given to the network, its Euclidean distance to all weight vectors is computed. The neuron whose weight vector is most similar to the input is called the winner. The weights of the winner and neurons close to it in the SOM are adjusted towards the input vector. The magnitude of the change decreases with time and with distance from the winner. This process is repeated for each input vector for a large number of cycles. An illustration of the training of a SOM is given in Figure 7.6. The gray cloud is the distribution of the training data, and the small red circle is the current training sample. At the beginning the SOM nodes are arbitrarily positioned in the data space. The node (highlighted in blue) which is nearest to the training sample is selected. It is moved towards the training sample, as are its neighbors on the grid. After many iterations the grid tends to approximate the data distribution (highlighted in green).

SOM for retinal image registration The theory of Kohonen's SOM can be adapted to the establishment of point correspondences between two images [Markaki et al., 2009]. In particular, the set of interest points from the reference image are considered as neurons of a neural network. Each weight vector holds the parameters of a local transformation. Each transformation maps an interest point and its neighborhood in the reference image to its correspondence in the second image. The parameters of the local transformations are calculated by means of an iterative optimization procedure that corresponds to the training of a neural network. At each step of the iterative procedure, a candidate perturbation (input vector) is randomly generated. This perturbation is used to update the transformation parameters of each point-neuron in analogy to Kohonen's SOM, taking also into account the spatial distribution of the points and their interactions. The update of the transformation parameters aims at optimizing of a measure of similarity between patches, centered at the points in the reference image space, and their transformed versions in the second image.

Angio2SLIM with SOM-LBP Before proceeding to the description of our approach it is important to note that it mostly follows the steps proposed by [Markaki et al., 2009] except for number of modifications. Therefore, we follow the same notation and provide certain description which can be found in the original work.

Let $\mu_A(I)$ denote the restriction of the image I to the region $A \subset \mathbb{R}^2$ and $T_{\mathbf{w}}(A)$ be the transformation with parameters $\mathbf{w} = [w_1, w_2, ..., w_K]$ of the region A, where K is the number of parameters needed for the definition of the specific transformation T. Given a pair of reference and target images, I^R and I^F respectively, a set of landmark points in the reference image $\{\mathbf{p}_i\}$ corresponds to the

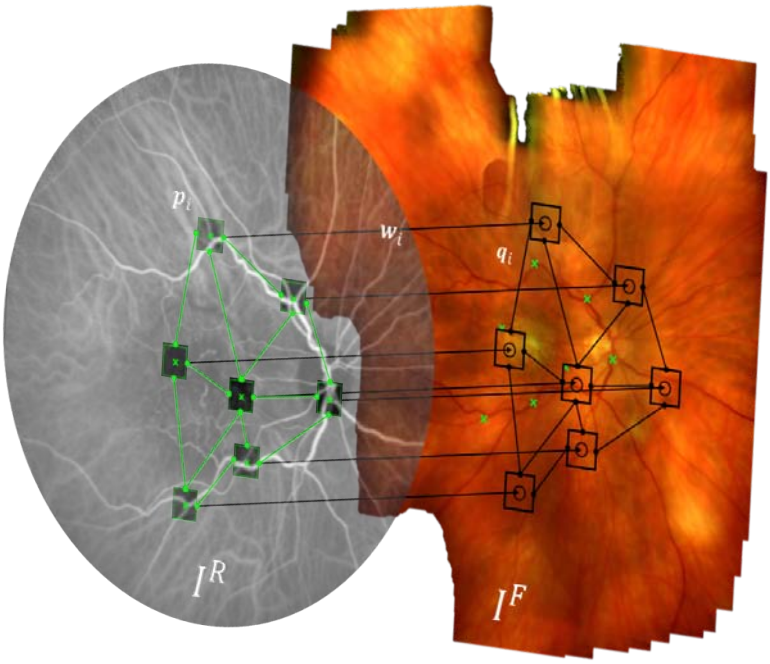


Figure 7.7: Schematic overview of the automatic point matching with SOM and LBP.

neurons in the network. Each point $\mathbf{p}_i = (x_i, y_i), i = 1, 2, ..., N$ is associated with a square image patch $A_i(n) = [x_i - r(n), x_i + r(n)] \times [y_i - r(n), y_i + r(n)]$, centered at it, where $r(n)$ is the side length (in pixels) of $A_i(n)$. The side length of $A_i(n)$ varies with the iteration number n of the network training. In particular, $r(n)$ is subject to an exponential decay from an initial value R_0 to a final value R_f during the training process, as described by the equation:

$$r(n) = \lfloor R_f + (R_0 - R_f)\varepsilon^{-c_r(n/n_{max})} \rfloor \tag{7.1}$$

where c_r is a decay constant, n_{max} the maximum number of iterations and $\lfloor\rfloor$ is the floor function.

Additionally, the weight vector $\mathbf{w}_i$, which holds the parameters of a local transformation $T_{\mathbf{w}_i}$ can be interpreted as a pointer to the target image I^F whose coordinates are the input to the network. This is schematically explained in Figure 7.7. Because $T_{\mathbf{w}_i}$ represents the local transformation which maps the region $A_i(n)$ of the reference image I^R to the region $T_{\mathbf{w}_i}(A_i(n))$ of the corresponding image I^F, an appropriate type of transformation, which can be best applied to the specific images, should be selected. In [Markaki et al., 2009] a similarity transformation was used. In our approach, however, we opted for the affine transformation and so $\mathbf{w}_i \in \mathbb{R}^6$ because it is sufficient for matching as it only needs to be valid locally. To quantify the correspondence between the square patch $\mu_{A_i(n)}(I^R)$ centered at neuron i and its corresponding region $\mu_{T_{\mathbf{w}_i}(A_i(n))}(I^F)$ a similarity metric $\mathfrak{M}_i$ is evaluated and assigned.

As mentioned in Section §7.1, the method in [Matsopoulos et al., 2004, Markaki et al., 2009] does not perform well on our dataset. Specifically, the $\mathfrak{M}$ based on gradient difference is not suitable in our case. The NMI, a popular choice in multi-modal registration, fails to take neighborhood relationships into account and does not improve on our data either. Thus, we introduce a different criterion to quantify correspondences between FA and SLIM that incorporates texture, spatial information, rotation and illumination invariance to $\mathfrak{M}$. This is achieved by extracting the LBP [Ojala et al., 2002] histograms from square image regions $\mu_{A_i(n)}(I^R)$ and $\mu_{T_{\mathbf{w}_i}(A_i(n))}(I^F)$ respectively. The resulting histograms are normalized so that entries sum up to 1. This is performed using the L_1-norm. It is worth mentioning that the 256-bit binary descriptor LBP can perform as well as the 128-dimensional float-

point-type SIFT but with a speed of more than two orders faster than SIFT while providing equal and sometimes even better discrimination ability [Ojala et al., 2002].

The training of the network is an iterative procedure, during which the optimal values of the parameters of each local transformation are determined. Before training begins, the n is set to 1, the weights of each neuron $\mathbf{w}_i$ are initialized to the parameters of the identity transformation and $\mathfrak{M}_i$ is calculated for the initial weights, according to :

$$\mathfrak{M} = \frac{1}{2}\sum_{b=1}^{B}\frac{(\mathbf{H}_b^R - \mathbf{H}_b^F)^2}{(\mathbf{H}_b^R + \mathbf{H}_b^F)} \tag{7.2}$$

where $b = 1,2,...,B$ is the number of histogram bins. The corresponding LBP histograms are calculated as $\mathbf{H}^R = LBP(\mu_{A_i(n)}(I^R))$ and $\mathbf{H}^F = LBP(\mu_{T_{\mathbf{w}_i}(A_i(n))}(I^F))$, where $LBP(A)$ is the local operator defined on image region A. Equation (7.2) is a Chi-squared distance that is smaller when histograms are more similar. The weight and the similarity measure at iteration n are denoted as $\mathbf{w}_i(n)$ and $\mathfrak{M}_i(n)$ respectively. Initially, $\mathbf{w}_i(n) \leftarrow \mathbf{w}_i$. During the training procedure, $\mathfrak{M}_i$ holds the lowest error found so far for a neuron i and $\mathbf{w}_i$ holds the corresponding weight vector. At every iteration n the following steps are performed:

1. A candidate perturbation of the current weight $\mathbf{dw}_i(n) = [dw_1(n), dw_1(n), ..., dw_K(n)]$ is randomly generated from the input space ζ, defined by setting the limits on every parameter of the affine transformation. Thus, every component of the perturbation vector is computed following the fast simulated annealing method [Ingber and Rosen, 1992] and falls within $[L_k, U_k]$ limits defined for each component, where $k = 1,2,...,K$. When a generated perturbation in not in the allowed range, then it is discarded and a new signal is produced until it satisfies the limits. The process controls how far from the currently best weights the perturbation can reach. As the iteration variable evolves, the generated input perturbations become more localized around the weights of the current winning neuron.

2. The perturbation is used to calculate the current similarity as:

$$\mathfrak{M}_i(n) = \mathfrak{M}(\mu_{A_i(n)}(I^R), \mu_{T_{\mathbf{w}_i+\mathbf{dw}_i(n)}(A_i(n))}(I^F)) \tag{7.3}$$

3. An update for the current weight vector for each point is calculated using the previously generated perturbation and the current similarity measure as:

$$\mathbf{w}_i(n) = \mathbf{w}_i + \alpha(\mathfrak{M}_i(n))\mathbf{dw}_i(n) + (1-\alpha(\mathfrak{M}_i(n)))\frac{\sum_j \mathfrak{M}_j G(i,j)(1-H(d))[\mathbf{w}_j - \mathbf{w}_i(n)]}{\sum_j \mathfrak{M}_j G(i,j)(1-H(d))} \tag{7.4}$$

$$\alpha(\mathfrak{M}_i(n)) = \frac{1}{1+e^{-s(c-\mathfrak{M}_i(n))}} \tag{7.5}$$

$$G(i,j) = \begin{cases} 0, & \| \mathbf{p}_i - \mathbf{p}_j \| \geq 3\sigma \\ e^{\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{2\sigma^2}} & otherwise \end{cases} \tag{7.6}$$

$$H(d) = \begin{cases} 0, & d < 0 \\ 1 & d \geq 0 \end{cases} \quad d = \mathfrak{M}_j - \mathfrak{M}_i \tag{7.7}$$

The update of the weight vector assigned to the current neuron depends on two factors, the random perturbation of weights presented to the network at the current iteration and the interaction between current neuron and its neighbors. Equation (7.5) is the sigmoid 'activation' function that determines the extent to which each of the two factors contributes to the weight

update. The parameter c is a threshold, above which the similarity $\mathfrak{M}_i(n)$ is considered "insufficient" and s is the slope of the function. Hence, the similarity value which does not exceed this threshold indicates that the pertubation induces a successful fitting (*i.e* the update is defined primarily by the pertubation) while the value above threshold shows the opposite and the update relies mostly on the interaction of the neighbor neurons. The contribution of the neighbor neurons (expressed by the fractional term in equation (7.4)) is based on two constraints, the neighbor neurons should fall within a certain distance from the current neuron i and their similarity values should be better than the best similarity $\mathfrak{M}_i$ of the neuron i found so far. The first condition is controlled by the equation (7.6), where $G(\cdot)$ is the Gaussian neighborhood function of neuron i with a standard deviation σ. It is evident that neighbor neurons located further than 3σ distance from neuron i will not be taken into account. The second condition is governed by the Heaviside step function $H(\cdot)$ given in equation (7.7), where d is the difference between similarity values of the neighbor neuron j and the current neuron i. Thus, the expression $(1-H(d))$ in the fractional term of the equation (7.4) ensures that only a neighbor with a better similarity value has an influence on the update for $\mathbf{w}_i(n)$. The contribution of each neighbor neuron to the update is normalized by the weighted sum of the similarity values of all neighbor neurons, expressed by the denominator of the fraction. If no neurons exist that satisfy these two constraints, then the fractional term of equation (7.4) is neglected.

4. The current similarity values are recomputed for the updated weights following equation (7.3) and compared with the best value $\mathfrak{M}_i$ found so far. Thus, if $\mathfrak{M}_i(n) < \mathfrak{M}_i$, the current weights $\mathbf{w}_i(n)$, as well as $\mathfrak{M}_i(n)$ are stored as best weights $\mathbf{w}_i$ and best similarity value $\mathfrak{M}_i$.

5. The training continues from step 1 until the maximum number of iterations has been reached or the convergence criterion has been satisfied. We define convergence as

$$\bar{\mathfrak{M}} < 10^{-5}, \quad \bar{\mathfrak{M}} = \frac{1}{n_0+1} \sum_{j=n-n_0}^{n} \bar{\mathfrak{M}}(j) \tag{7.8}$$

where $\bar{\mathfrak{M}}(j)$ is the average value of the best similarity of the neurons at the j-th iteration and n_0 is a predefined number of iterations.

In the end of the training procedure each neuron will correspond to the landmark detected on the FA image and the best $\mathfrak{M}_i$ of node i provides an estimation of the quality of the obtained correspondence. Usually SLIM provides a wider coverage of the retina compared to FA. Additionally, even if SLIM is constructed precisely and the amount of drift is minimized, the chance of having misalignment still exists. Both these factors may eventually lead to outliers in correspondences obtained by SOM-LBP. For example, a bifurcation point detected in the FA image may not be present in the SLIM image and the algorithm will then find the most similar false positive correspondence. To detect such cases we benefit from the estimated quality of each correspondence by means of $\mathfrak{M}_i$ and simply discard the points that do not satisfy $\mathfrak{M}_i < th$. In our experiments we found that a good value for the threshold is $th = 0.7$. After rejection of the outliers, the best weights define the local transformation which maps the point $\mathbf{p}_i$ from the reference image I^R to the point $\mathbf{q}_i = T_{\mathbf{w}_i}(\mathbf{p}_i)$ on the target image I^F.

7.2.3 Non-rigid image registration with the normalized quadratic model

Once point correspondences $\mathbf{p}_i \longleftrightarrow \mathbf{q}_i$ are estimated we invoke the model fitting procedure where the transformation parameters are calculated. It was shown in Chapter 4 that an affine transformation model is sufficient for mono-modal SLIM registration. However, in the case of multi-modal registration a model of higher complexly is needed to account for the significant geometrical differences between modalities and ensure an accurate mapping of the geometry of the human eye. We, thus, use

the quadratic model, which was specifically derived to fit the curved retinal surface [Can et al., 2002]. It is a second order Taylor series expansion of the general quadratic transformation

$$Q(\mathbf{p}) = \left[\mathtt{B}_{2\times3}|\mathtt{A}_{2\times2}|\mathbf{t}_{2\times1}\right] X(\mathbf{p}) \tag{7.9}$$

where $\mathtt{B} \in \mathbb{R}^{2\times3}$, $\mathtt{A} \in \mathbb{R}^{2\times2}$, $\mathbf{t} \in \mathbb{R}^{2\times1}$ are the 2nd, 1st and 0th order terms of the transformation, and $X(\mathbf{p}) = [x^2, xy, y^2, x, y, 1]^\top$. We use a normalized estimation procedure proposed specifically for the quadratic model to avoid numerical instability induced by the squared terms. This includes computing a normalization N and a denormalization D' transforms from $\{\mathbf{p}_i\} \longleftrightarrow \{\mathbf{q}_i\}$ respectively. We refer the reader to Chapter 4 and specifically to §4.3.3 of this manuscript for more details. Fitting is then performed by means of least squares to obtain $\tilde{Q}$ and the final normalized quadratic transformation Q is calculated using $(D' \circ Q)(\mathbf{p})$.

7.3 Experimental results and discussion

The goal of the evaluation is two-fold. First, to verify and demonstrate the significance of the inclusion of the priority landmark detection into the registration method. Second, to compare the registration accuracy of the proposed method with the baseline method and with the current state-of-the-art. The description of the datasets is given in §7.3.1 along with the details on ground truth and evaluation criteria. The experiments and discussion are given in §7.3.2 - §7.3.4.

7.3.1 Datasets and ground truth

The proposed registration model was evaluated on three *in-vivo* datasets including multi-modal and mono-modal data:

- *FA2SLIM* (multi-modal): a dataset of 20 image pairs from 11 patients associated with unknown transformations was obtained from 11 different patients presented with healthy and unhealthy retinas at University Hospital of Saint-Etienne, France. The NPRP system developed at Quantel Medical was used to obtain slit-lamp video sequences. SLIM was created using the mosaicing technique described in Chapter 5. The corresponding FA images were obtained using the Diagnostic Imaging Platform Heidelberg SPECTRALIS®. This dataset is relatively small. This is because SLIM is a recent retinal image modality acquired with an experimental industrial prototype. Its usage is not always complemented by sessions where the corresponding FA images can be obtained.

- *FA2Fundus* (multi-modal): a database of 60 pairs of FA and corresponding color fundus (CF) images of 30 healthy persons and 30 patients with diabetic retinopathy [Shirin et al., 2012]. The images were obtains with a fundus camera using excitation and barrier filters for FA data.

- *FIRE* (mono-modal): a dataset provided by [Hernandez-Matas et al., 2016] with 134 fundus image pairs from 39 patients. The images were acquired with a Nidek AFC-210 fundus camera, which acquires images with a resolution of 2912×2912 pixels and a FOV of $45°$ both in the x and y dimensions.

In retinal *in-vivo* data, a ground truth for image registration is hard to obtain due to numerous factors such as the exact size of the human eye, the distortions caused by cornea and the tiny movements of the eye. Moreover, in the case of SLIM, the optical set-up and the image composition bring additional complications. Thus, to generate the GT data for *FA2SLIM* dataset we manually selected a set of 10 landmarks, *i.e.*, bifurcations and corner points, from the original images, denoted $\bar{\mathbf{p}}_j \longleftrightarrow \bar{\mathbf{q}}_j, j = 1, 2, ..., M$. This allowed us to account for poor quality images and images presenting

an unhealthy retina. Moreover, this has a benefit of providing reliable and fair measurements over all images in a dataset. Because the *FA2Fundus* dataset does not contain annotations, we also generated the GT point correspondences manually. This was not the case with *FIRE* as the authors provide the GT annotations. To adapt the *FA2Fundus* and *FIRE* datasets to our experiments we discarded out the image pairs where the OD is not visible. This is necessary to ensure that the macula detection step works as it depends on the OD localization. All datasets consist of both normal and abnormal retinal images. The abnormal retinal images exhibit visual anatomical differences, due to the progression or remission of retinopathy. These differences may appear in the form of increased vessel tortuosity, microaneurysms, cotton-wool spots, etc. Sample image pairs from each dataset are shown in Figure 7.8 and a summary with corresponding GT data is given in Table 7.1.

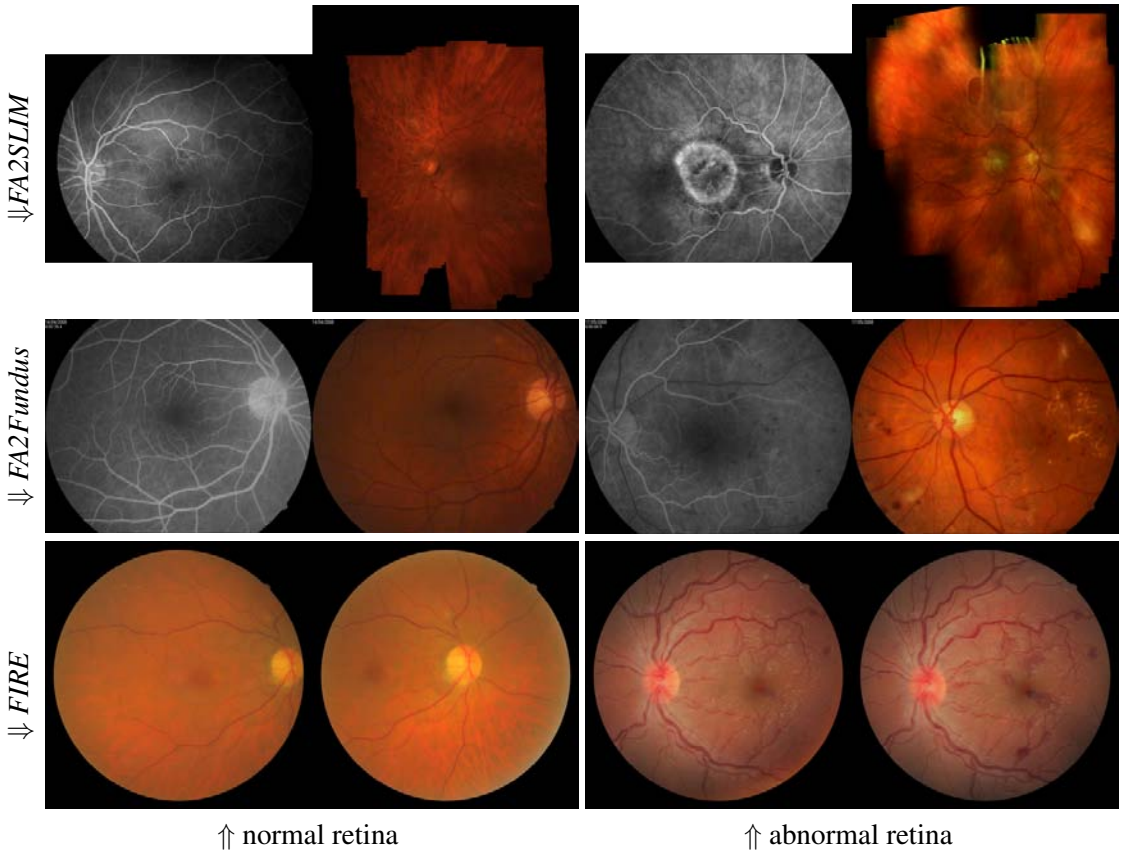


Figure 7.8: Sample image pairs from the different datasets. The first column shows the normal retinal image pairs to be registered while the second column illustrates the abnormal cases within a corresponding dataset, indicated row-wise.

$I_{O^{any}}$ **# image pairs**	20	60	134
	FA2SLIM	*FA2Fundus*	*FIRE*
$I_{O^{vis}}$ **# image pairs**	20	56	129
Normal	50%	48%	89%
Abnormal	50%	52%	11%
GT (# points)	10	10	10

Table 7.1: The characteristics of the used datasets and corresponding ground truth. $I_{O^{any}}$ denotes the initial number of images where the OD was either visible or not. $I_{O^{vis}}$ denotes the number of images where the OD is fully visible. Subsequent Normal, Abnormal and GT values correspond to $I_{O^{vis}}$.

7.3.2 Inclusion of the priority landmark detection

The vessel bifurcation point extraction is a combination of relatively standard image processing techniques and it can be replaced by any existing method with comparable results. Thus, we prefer to assess the performance of our landmark detection stage from the perspective of inclusion of the macula localization step. The obvious impact on the registration step is the enlargement of the point set used for registration. Even though the second stage of our approach - automatic point correspondences - does not require a large point set to establish good point correspondences, the non-rigid registration with quadratic model fitting demands more than 20 point correspondences to provide acceptable results.

To quantitatively evaluate the effect of the landmark detection step on the registration result we obtain point correspondences by the proposed SOM-LBP with and without points from the macula detection step. We use them to estimate the normalized quadratic transformation as Q^+ and Q^-. We then compute the RMSE between the reprojected GT points from the reference image and their corresponding points in the target image as:

$$\mathscr{E}^{+/-}_{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(\bar{\mathbf{q}}_j - Q^{+/-}(\bar{\mathbf{p}}_j)^2} \tag{7.10}$$

The RMSE computed for 10 randomly chosen image pairs from each dataset is shown in Table 7.2.

		S01	**S02**	**S03**	**S04**	**S05**	**S06**	**S07**	**S08**	**S09**	**S10**	**AVG**
FA2SLIM	N^+	43	27	44	31	48	26	27	31	34	26	**34.3**
	$\mathscr{E}^+_{RMSE}$	1.31	2.48	4.22	1.25	2.30	1.44	2.33	3.21	3.27	4.35	**3.28**
	N^-	40	23	40	26	40	19	21	30	28	22	**28.9**
	$\mathscr{E}^-_{RMSE}$	1.35	2.56	4.23	1.28	2.34	1.51	2.35	3.20	3.28	4.36	**3.37**
FA2Fundus	N^+	26	40	44	31	45	34	46	27	33	37	**35.1**
	$\mathscr{E}^+_{RMSE}$	0.28	0.20	0.21	0.34	0.23	0.29	0.21	0.31	0.24	0.28	**0.26**
	N^-	22	35	37	25	41	32	42	21	28	32	**29.6**
	$\mathscr{E}^-_{RMSE}$	0.36	0.29	0.28	0.42	0.24	0.31	0.23	0.40	0.28	0.30	**0.35**
FIRE	N^+	40	24	45	42	38	41	34	35	39	43	**38.1**
	$\mathscr{E}^+_{RMSE}$	0.23	0.30	0.19	0.22	0.26	0.21	0.28	0.27	0.25	0.21	**0.22**
	N^-	37	22	39	36	34	37	32	29	35	37	**33.8**
	$\mathscr{E}^-_{RMSE}$	0.26	0.36	0.23	0.27	0.29	0.27	0.31	0.30	0.29	0.28	**0.27**

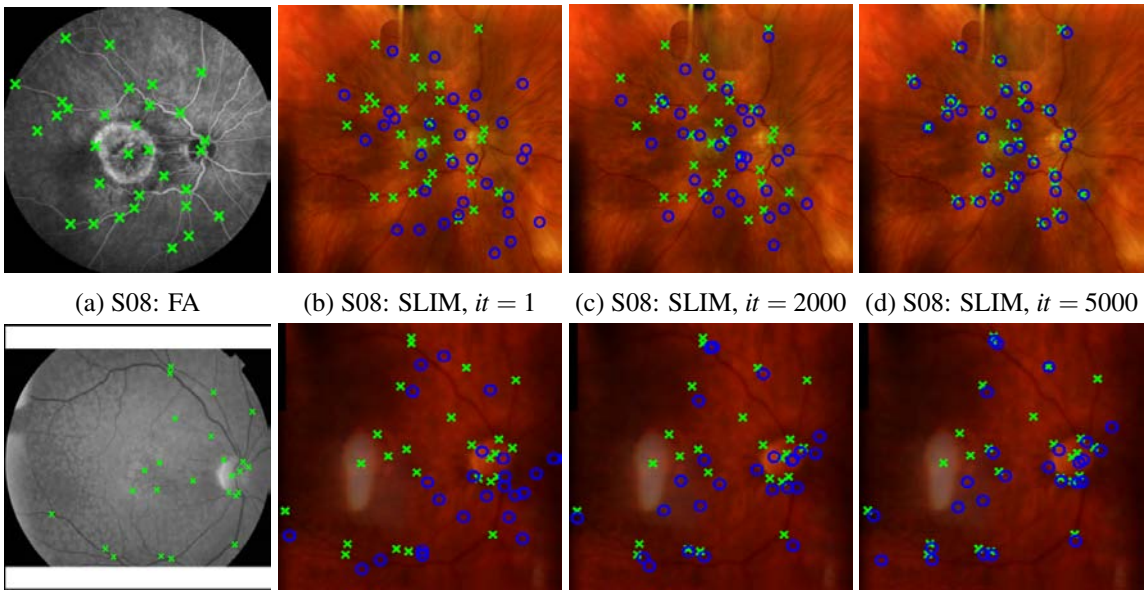
Table 7.2: Evaluation of the effect of the inclusion and exclusion of the macula detection step on the registration errors.

The number of points with and without macula detection step is noted as $N^{+/-}$ while the corresponding errors as $\mathscr{E}^{+/-}_{RMSE}$. The **AVG** in the last column stands for the results averaged over all images from the corresponding datasets. In order to ensure the consistency, all the numbers were averaged over 5 independent executions of the algorithm. For the *FA2SLIM* dataset, the number of landmark points in the reference image, extracted without inclusion of the macula detection step varies from 19 to 40 points for the 10 samples shown in the table. The inclusion of this step brings between 1 and 9 new points. The average number of points over the whole dataset without the macula detection is 28.9 which is increased to 34.3 by including the step. At first sight, a gain of 5 points does not seem to be significant. The registration errors, however, give a different perspective. As one can see, the registration error for the 10 representative samples always improves when more points are added. This, however, does not correlates with the number of points but rather with their quality. For example,

one can see that 4 points added in **S02** have reduced the error by 0.8 pixels while 8 points added in **S05** have reduced the error only by 0.4 pixels. In samples **S09** and **S10**, the errors were reduced only by 0.1 pixels with the inclusion of 6 and 4 macula points. A very similar tendency can be observed in the statistics for the *FA2Fundus* and *FIRE* datasets. Thus, one can see, that the registration error improves when more points are added.

7.3.3 Automatic point matching

The visual evaluation of the proposed SOM-LBP algorithm for automatic point matching is shown in Figure 7.9. The initial location of the points (before training) is shown in the first column. This has been achieved using pre-alignment with respect to the center of the SLIM image. The subsequent columns show the evolution of the algorithm through the training in a zoom-in region of the target image. The algorithm usually converges to an optimal solution after 5000 iterations. Second row illustrated an example with uncorrected mismatches. This is due to the fact that points detected on the FA image were not present on the SLIM image.



(a) S08: FA (b) S08: SLIM, $it = 1$ (c) S08: SLIM, $it = 2000$ (d) S08: SLIM, $it = 5000$

(e) S10: FA (f) S10: SLIM, $it = 1$ (g) S10: SLIM, $it = 2000$ (h) S10: SLIM, $it = 5000$

Figure 7.9: Performance of the automatic point matching algorithm between an AF and SLIM sample image pairs from the *FA2SLIM* dataset. Green crosses show the GT point correspondences while blue circles indicate the detected points using the proposed SOM-LBP method. An example with mismatched points is shown in the second row.

To evaluate the quality of our automatic point matching quantitatively we compute a percentage of correctly match keypoints, denoted as $\mathfrak{M}_{PCK}$. Let $\bar{\mathbf{p}}_i \longleftrightarrow \tilde{\mathbf{q}}_i$ be the point correspondences found by the proposed SOM-LBP method, where $\bar{\mathbf{p}}_i$ is given as an input. Then, $\tilde{\mathbf{q}}_i$ is considered to be matched correctly if it falls within a local neighborhood around $\bar{\mathbf{q}}_i$ (*i.e.*the GT point that corresponds to $\bar{\mathbf{p}}_i$) such as $\tilde{\mathbf{q}}_i \in V(\bar{\mathbf{q}}_i)$, where $V(\bar{\mathbf{q}}_i) = \{\tilde{\mathbf{q}} | \parallel \tilde{\mathbf{q}} - \bar{\mathbf{q}}_i \parallel \leqslant d\}$ with $d = 5$ pixels. We provide consolidated statistics in Table 7.3, where the mean $\mathfrak{M}_{PCK}$ over datasets is shown for comparison. One can see that the mean $\mathfrak{M}_{PCK}$ for our *FA2SLIM* dataset is the lowest one. This is because the FA and SLIM image modalities are fundamentally different compared to the images from the *FA2Fundus* dataset where the same fundus camera was used for color fundus images and FA was obtained by applying excitation and barrier filters. The mono-modal dataset *FIRE*, as was expected, provides the highest values, especially because it has the lowest number of abnormal examples that can affect the performance. A

similar pattern can be observed when looking at the statistics on normal and abnormal cases. Overall it can be concluded that the mosaic quality has a great impact on the proportion of correct matches in the *FA2SLIM* dataset. This confirms that multi-modal registration with SLIM is a difficult task.

	FA2SLIM	*FA2Fundus*	*FIRE*
Mean $\mathfrak{M}_{PCK}$ over full dataset	89%	92%	97%
Mean $\mathfrak{M}_{PCK}$ over *normal* cases	84%	88%	95%
Mean $\mathfrak{M}_{PCK}$ over *abnormal* cases	73%	81%	83%

Table 7.3: Accumulated statistics over three datasets on Proportion of Correctly matched Keypoints $\mathfrak{M}_{PCK}$ obtained with SOM-LBP.

7.3.4 Multi-modal registration: comparative results

Evaluating the accuracy of retinal image registration is not an easy task because of the lack of ground truth. In this experiment we compute two types of errors: RMSE and the CEM, to compare the performance of the proposed method across datasets and with the baseline and the state-of-the-art methods on our dataset. We consider a semiautomatic registration method currently used at QuantelMedical as the baseline. This involves manual selection of point correspondences by a human operator followed by an automatic estimation of the rigid transformation from the selected points, which is then used to register the FA image to the SLIM image. We also compare the proposed method with the GDB-ICP method proposed by [Yang et al., 2007], the original SOM-based registration by [Markaki et al., 2009] that we used as the basis for our framework and the method based on specifically designed PIIFD retinal feature descriptor presented by [Chen et al., 2010]. In addition, the success rate was investigated in our experiments as well. In our settings, a registration was considered successful if the RMSE was at least 50% lower compared to the baseline. It is important to note that the computation of the successful rate for the baseline method is not applicable because it is a semiautomatic method where the point correspondences were selected manually. The RMSE is calculated implying that the macula detection is performed by default. We, thus, note it as $\mathscr{E}_{RMSE}$ without superscript. About 80% of reprojected points would be expected to lie within GT points on the successfully registered images. This is computed as

$$\mathscr{E}_{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(\bar{\mathbf{q}}_j - \mathscr{T}(\bar{\mathbf{p}}_j, \theta))^2} \tag{7.11}$$

where $\mathscr{T}$ is the transformation function with parameters θ estimated by a method under comparison. The CEM is the second type of error in our evaluation. It is defined as median distance over a set of centerline point correspondences C located on a major axes centerline between two registered images

$$\mathscr{E}_{CEM} = \underset{(\bar{\mathbf{p}}_j, \tilde{\mathbf{q}}_j)\in C}{\text{median}} |\bar{\mathbf{q}}_j - \mathscr{T}(\bar{\mathbf{p}}_j, \theta)| \tag{7.12}$$

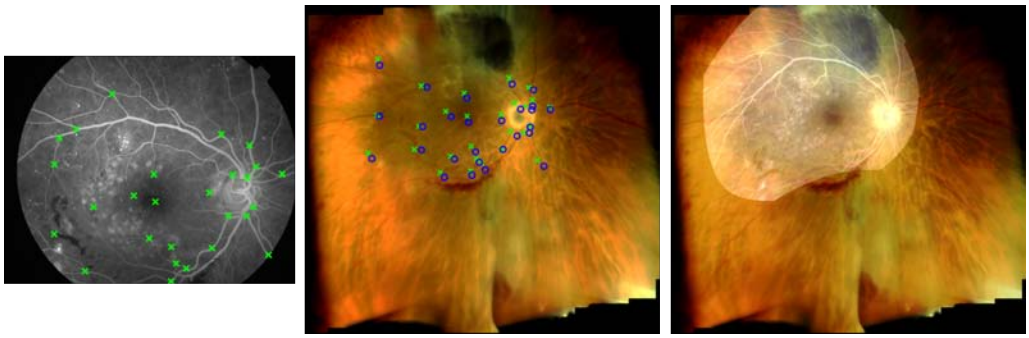
Evaluation across different datasets We provide the mean error statistics in pixels (px) for all the images across three different datasets in Table 7.4. One can see that the results on the *FIRE* dataset provide the lowest errors. This is because it is a mono-modal case. The algorithm, however, achieve a subpixel accuracy only for image pairs of healthy retina (0.94px and 0.53px) for this dataset. A slightly different picture appear for the performance on the *FA2Fundus* dataset. Here, the overall mean does not exceed 2 pixels (1.56px and 1.83px) and is close to 1 pixel for the normal retina cases (1.14px and 1.15px). The performance of our algorithm on abnormal cases across all datasets is

slightly inferior compared to the normal cases (the rest of the images), which may be attributed to a minimal number of good point correspondences due to the presence of the retinal abnormalities in the regions of interest. In our *FA2SLIM* dataset the errors go beyond 2 pixels (2.27px and 2.16px) and even more for the abnormal retina cases ($\mathscr{E}_{RMSE} = 3.10$px). These results illustrate the proposed approach has a potential to generalize to other retinal image modalities and that it can be successfully applied to a mono-modal registration. The registration results on 4 sample pairs from our *FA2SLIM* dataset are shown in Figure 7.10 for visual assessment with the corresponding errors.

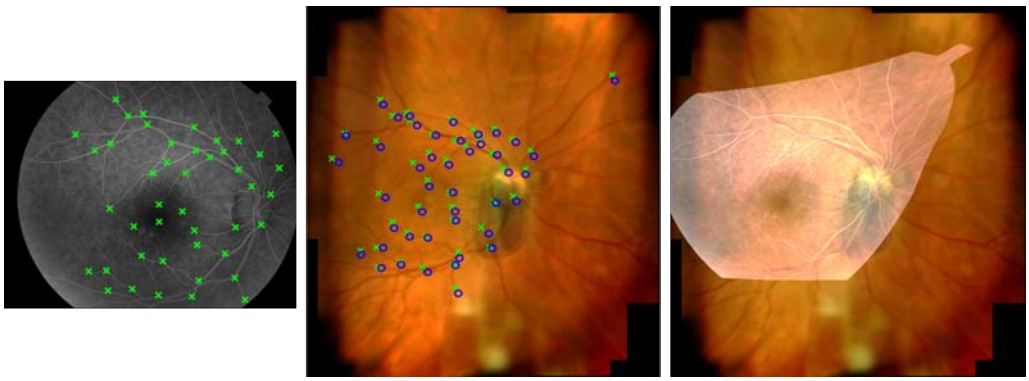
	FA2SLIM		*FA2Fundus*		*FIRE*	
	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$
Mean over full dataset	2.27	2.16	1.56	1.83	1.02	1.41
Mean over *normal* cases	1.44	1.85	1.14	1.15	0.94	0.53
Mean over *abnormal* cases	3.10	2.48	2.01	1.92	1.94	1.86

Table 7.4: Mean registration error statistics for the proposed method across three datasets.

(a) Pair S03 : $\mathscr{E}_{RMSE} = 4.22 \mid \mathscr{E}_{CEM} = 3.72$



(b) Pair S07 : $\mathscr{E}_{RMSE} = 2.33 \mid \mathscr{E}_{CEM} = 2.80$



(c) Pair S15 : $\mathscr{E}_{RMSE} = 3.85 \mid \mathscr{E}_{CEM} = 3.18$

(d) Pair S08 : $\mathscr{E}_{RMSE} = 3.21 \mid \mathscr{E}_{CEM} = 1.12$

Figure 7.10: Image registration results on sample image pairs from the *FA2SLIM* dataset. From left to right the columns show the reference image with detected landmarks, the target image with detected correspondences (blue) and their GT points (green) and the registration result as a fused image.

Comparison with the state-of-the-art GDB-ICP is a generalized version of DB-ICP, which was originally proposed for retinal image registration. We obtained a public copy of its binary program and tested it with the quadratic transformation parameter setting. We implemented the method by [Markaki et al., 2009] (SOM-09) in Matlab and also found the Matlab implementation of the PIIFD on the Internet. Quantitative results from the application of the proposed, the baseline and the state-of-the-art methods for the 20 image pairs of our dataset *FA2SLIM* are listed in Table 7.5.

	Angio2SLIM		**SOM-09**		**GDB-ICP**		**PIIFD**		**Baseline**	
	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$	$\mathscr{E}_{RMSE}$	$\mathscr{E}_{CEM}$
pair S01	1.31	3.72	1.98	4.39	-	-	3.86	6.27	5.38	7.79
pair S02	2.48	2.15	3.89	3.55	-	-	-	-	8.57	8.24
pair S03	4.22	3.65	5.63	5.06	-	-	-	-	12.18	11.61
pair S04	1.25	1.32	2.18	2.25	-	-	2.56	2.63	5.22	5.29
pair S05	2.30	1.18	3.81	2.69	-	-	4.02	2.90	9.62	8.50
pair S06	1.44	1.42	2.52	2.50	-	-	-	-	5.81	5.79
pair S07	2.33	2.80	3.17	3.64	-	-	-	-	7.38	7.85
pair S08	3.21	1.12	4.87	2.78	-	-	-	-	10.04	7.95
pair S09	3.27	2.64	4.64	4.01	-	-	-	-	8.20	7.57
pair S10	4.35	3.71	5.71	5.07	-	-	-	-	10.12	9.48
pair S11	1.36	1.38	2.98	3.00	-	-	4.64	4.66	5.15	5.17
pair S12	1.05	1.12	1.76	1.83	-	-	3.28	3.35	5.26	5.33
pair S13	2.34	3.01	3.02	3.69	-	-	4.85	5.52	4.87	5.54
pair S14	3.12	2.98	4.89	4.75	-	-	-	-	7.32	7.18
pair S15	3.85	3.18	4.48	3.81	-	-	-	-	6.75	6.08
pair S16	1.12	1.00	1.74	1.62	2.86	2.55	2.38	2.11	4.36	4.24
pair S17	2.08	2.56	2.95	3.43	-	-	-	-	5.58	6.06
pair S18	1.91	1.45	2.52	2.06	-	-	-	-	8.11	7.65
pair S19	1.35	1.13	1.76	1.54	-	-	3.12	2.9	5.61	5.39
pair S20	1.18	1.86	2.05	2.73	-	-	3.89	4.57	5.13	5.81
MEAN	**2.27**	**2.16**	**3.32**	**3.22**	**n/a**	**n/a**	**n/a**	**n/a**	**7.03**	**6.92**
MEDIAN	**2.19**	**2.00**	**3.00**	**3.21**	**n/a**	**n/a**	**n/a**	**n/a**	**6.28**	**6.63**
STD	**1.07**	**0.98**	**1.32**	**1.09**	**n/a**	**n/a**	**n/a**	**n/a**	**2.18**	**1.77**
Success rate	**85%**		**60%**		**5%**		**20%**		**n/a**	

Table 7.5: Quantitative results of two error metrics, $\mathscr{E}_{RMSE}$ and $\mathscr{E}_{CEM}$, obtained by the proposed method (Angio2SLIM), the semiautomatic method used in TrackScan (Baseline) and the existing algorithms proposed in [Yang et al., 2007] (GDB-ICP), [Markaki et al., 2009] (SOM-09) and [Chen et al., 2010] (PIIFD) for image pairs of the *FA2SLIM* dataset (errors in pixels).

The failure to register an image pair is indicated with a dash symbol. One can see that GDB-ICP failed to provide registration result on the majority of image pairs. Only one image pair (S16) was successfully registered. This is because in that particular example no retinal abnormalities were present, the mosaic had almost no illumination artifacts, neither it had holes in it and it was almost the size of the corresponding FA image. This led to reliable correspondences that were evenly distributed in almost the entire image. One can also notice, that the error statistics for this case for the other methods are the lowest compared to the other cases. A better job was done by the PIIFD method, which, however, failed to register more than 50% of the dataset (11 pairs). It is important to note that these images were all examples of the abnormal retina cases except one (S02). Therefore, it is impossible to compute the mean/median/std statistics, which are indicated as 'na' in the table. The mean errors of the Baseline method (7.03px and 6.92px) indicate the quantitative result that we want to improve on with our method. The corresponding values for the proposed approach Angio2SLIM

are close to 2 pixels (2.27px and 2.16px), where the median is 2.19 and 2 pixels respectively and the standard deviation is relatively low (1.07px and 0.98px). This certifies the reproducibility of our approach. Moreover, one can see that our algorithm outperforms the Baseline registration and the results provided by SOM-09. We improve over the Baseline registration in both error metrics for 5.03 and 4.76 pixels on average (2.27px vs. 7.03px and 2.16px vs. 6.92px respectively). A similar trend can be observed while comparing our Angio2SLIM with SOM-09 with a different order of magnitude, where the improvement in mean errors are 1.05px and 1.06px respectively (2.27px vs.3.32px and 2.16px vs.3.22px in Table 7.5). We gained a performance boost for over than 1 pixel in 9 image pairs. This may not sound impressive if one registers images obtained with a fundus camera. In case of SLIM, however, we consider this as reliable. Both algorithms, the proposed one and SOM-09, have not achieved a subpixel accuracy on our dataset. The proposed approach, however, successfully registered 90% of the image pairs, compared to SOM-9 which only succeeded for 60%. This once more verifies the significant performance lift achieved with the introduced inclusion of macular points and LBP-based training. Moreover, our method is fully automatic and satisfies the fundamental requirements of the TrackScan platform.

7.4 Conclusion

We have presented a new method for registering the FA and SLIM retinal image modalities without manual input. The method starts with the detection of important anatomical landmarks on the reference image, complemented by the localization of the priority region. The point correspondences on the target image are established in an unsupervised iterative stochastic optimization. A data driven LBP measure of correspondence quality is used in the process. Although, LBP has the advantage of tolerance of illumination changes and computational simplicity, it induces an ambiguity to the estimation of the affine parameters for SOM's weights in the sense that the rotation parameter does not have much impact. This, however, does not interfere with the global solution as the affine transform is used to align points locally while a higher order transformation is used afterwards for global registration. The final registration is achieved by fitting a normalized quadratic model to the point correspondences. Although our system builds on the existing methods, with the various specific modifications made, it has been shown to be effective for the aforementioned task and significantly improves the baseline registration method.

Chapter 8

Conclusions

In this chapter we analyze what has been done with respect to the initial objectives. We draw a number of conclusions in §8.1. We also reflect on the future research directions in §8.2.

Contents

8.1 Conclusion **100**

8.2 Future work **101**

8.1 Conclusion

Global estimates on visual impairment reported by the WHO show that the principal cause of blindness is cataract. However, retina related disorders such as DR and AMD, both referred as retinopathies, are the leading causes of preventable blindness among working populations in economically-developed societies. With special imaging and examination methods it is possible to perform direct *in vivo* non-invasive observation of the retina and identify the type and stage of the retinopathy. For more than 50 years, laser technology has evolved to become an essential tool in the treatment of diabetic retinopathies. Navigated laser pan-retinal photocoagulation is considered the standard treatment worldwide. While the fundus camera based systems are considered the optimal choice, the magnification and control offered by the conventional slit-lamp still makes it a very popular choice in the clinical environment. Moreover, the slit-lamp based systems, which date back to the 1980s, are the common technology used for this treatment that every ophthalmologist is familiar with. In this context the development of a platform to combine conventional slit-lamp laser delivery with computer-assisted navigation is on demand.

Recently, a computer assisted slit-lamp based industrial prototype TrackScan has been developed in QuantelMedical. The prototype combines real-time HD imaging, pre-operative planning and intra-operative navigation. It also provides the basic functionality for multi-modal registration of diagnostic images. The slit-lamp biomicroscope makes it possible to obtain a SLI with a high resolution which allows them to perform an accurate laser shot. The counterpart is that the visualization is local and does not help the practitioner in their therapeutic act. SLIM is used for view expansion and treatment planning. However, mosaicing slit-lamp images is a difficult task due to the absence of a physical model of the imaging process and mosaicing drift. Furthermore, the specifics of the imaging setup introduce bothersome illumination artifacts. They not only degrade the quality of the mosaic but may also affect the diagnosis. In this manuscript we presented our contributions to the problem of precise SLIM and automatic multi-modal registration of SLIM with FA to assist navigated pan-retinal photocoagulation. The main focus was to propose improvements in various aspects of the baseline SLIM method currently implemented in TrackScan. To this end we set up a number of objectives which we fulfilled in the course of this thesis project.

We conducted a comparative study of transformation models using a specifically designed feature-based evaluation framework. This allowed us to demonstrate that the quadratic transformation, widely used in retinal image registration, is unstable on our data even after improvement by applying the specifically derived normalization procedure. This led us to conclude that an affine model is the best compromise between ability to model pairwise transformations and simplicity in dealing with mosaicing drift in SLIM. The choice of an appropriate transformation was what we were looking for at the early stage of our research. The conclusion on the right transformation then defined the basis for our mosaicing method with drift reduction. We formulated our solution based on key-frame BA and proposed a local refinement procedure. We verified that point correspondences presented in multiple views provide more constraints. A simple global motion model associated with local correction is a valid assumption which guarantees the prediction of the track location. This helped us to obtain tracks longer than short-inter-frames with improved precision. We demonstrated that using a simple global model to initialize key-frame based local BA can be as accurate as performing global BA. The results that we obtained, showed an improvement over the baseline method implemented in the Trackscan as it was our primary objective.

Our second contribution was directed to the problem of specular highlights of various degrees and their effect on the photometric quality of the mosaics. We demonstrated that traditional approaches are not suitable for SLIM. Therefore, we proposed a better alternative by designing a method based on a fast single-image technique to remove glares. We have used the notion of the type of semi-transparent specular highlights and motion cues for intelligent correction of lens flares. This allowed

us to improve the rendering of the mosaics significantly and achieve a better visual results with less specular reflections and wider coverage compared to the solution currently implemented in Trackscan. In our last contribution we attempted to solve the problem of the multi-modal registration between the FA and SLIM. Overcoming the issue of the detection of point correspondences from both images, we formulated our solution as an unsupervised learning procedure with SOM that takes as input a set of detected landmarks from the FA image and establishes reliable point correspondences with SLIM via unsupervised training with self-organization. We complemented this by incorporating the detection of macula area to ensure an accurate registration in the priority landmark. Thus, the proposed method is the first that is able to register FA and SLIM without manual input by detecting anatomical landmarks only on one image and ensuring an accurate registration in the macular area.

Parts of the work presented in this thesis have led to the publication of articles in several peer-reviewed scientific journals and conferences

8.2 Future work

There are several possible directions for improvement. We list them below with respect to the initial objectives set for the thesis:

SLIM and mosaicing drift The current solution can possibly be improved by converting it to an incremental algorithm and incorporating the Simultaneous Localization and Mapping (SLAM) approach. Along with mosaicing, super-resolution techniques can be used to fuse the information across the view. This would help in generating better quality images with increase in pixel resolution.

Light-related imaging artifacts The current solution is only capable of removing glared regions from the retina in an on-line fashion while leaving the further corrections as the post-processing step. This can be possibly improved by transferring the post-processing step into a predictive photometric tracking.

Multi-modal registration of SLIM and FA Foremost is to extend the current solution to FA images where the optic disc is not fully visible. We can benefit from recent advances in the application of Deep Learning (DL) to the task of retina vessel segmentation. To this end, the existing pre-trained models can be used and fine-tuned on our dataset of FA images. Moreover, the notion of priority landmark can be incorporated in the model which may potentially boost the performance. The application of DL can also help to learn a data specific similarity metric and simultaneously estimate the parameters of the geometric transformation.

General aspects The proposed set of improvements to TrackScan is currently implemented in Matlab. By porting it into C++ along with parallelization would help in building a real-time system. Furthermore, a detailed clinical study of the proposed tools would help in understanding the impact in general usage and also required improvements.

Abbreviations

AMD	Age-related Macular Degeneration
BA	Bundle Adjustment
BCM	Blending Consistency Measure
BRIEF	Binary Robust Independent Elementary Features
CC	Cross Correlation
CEM	Centerline Error Measure
CF	Color Fundus Photography
CIFRE	Conventions Industrielles de Formation par la REcherche
CLAHE	Contrast Limited Adaptive Histogram Equalization
CT	Computed Tomography
DBICP	Dual-bootrsrap Iterative Closest Point
DL	Deep Learning
DOF	Degrees of Freedom
DR	Diabetic Retinopathy
DSC	Dice Similarity Coefficient
ECC	Enhanced Correlation Coefficient
EKF	Extended Kalman Filter
EM	Expectation Maximization
EnCoV	Endoscopy and Computer Vision
ESM	Efficient Second-Order Minimization
FA	Fluorescein Angiography
FBC	Forward-Backward Constancy
FLANN	Fast Library for Approximate Nearest Neighbors
fMRI	functional Magnetic Resonance Image
FOV	Field of View
FPGA	Field-programmable Gate Array

GA	Genetic Algorithm
GDBICP	Generalized Dual-bootrsrap Iterative Closest Point
GDm	Gradient Descent method
GF	Glare-Free
GMM	Gaussian Mixture Model
GPU	Graphics Processing Unit
GT	Ground Truth
GTM	Graph Transformation Matching
GUI	Graphical User Interface
HD	High Definition
ICP	Iterative Closest Point
IR	Infra-red
ISODATA	Iterative Self-Organizing Data Analysis Technique
KLT	Kanade-Lucas-Tomasi
LBP	Local Binary Patterns
LCE	Loop Closure Error
LDDMM	Large Deformation Diffeomorphic Metric Mapping
LFE	Local Fitting Error
LLS	Linear Least Squares
MAD	Mean Absolute Differences
MI	Mutual Information
minEig	Minimum Eigen Value algorithm
ML	Machine Learning
MRI	Magnetic Resonance Image
MSD	Mean Squared Difference
MST	Minimum Spanning Tree
NCC	Normalized Cross Correlation
NLLS	Non-linear Least Squares
NMI	Normalized Mutual Information
NN	Nearest Neighbor
NPDR	Non-proliferative Diabetic Retinopathy
NPRP	Navigated Pan-retinal Photocoagulation
OCT	Optical Coherence Tomography
OD	Optic Disc
PCA	Principal component Analysis
PDE	Partial Differential Equations
PDR	Proliferative Diabetic Retinopathy
PET	Positron Emission Tomography
PIIFD	Partial Intensity Invariant Feature Descriptor
PRP	Pan-retinal Photocoagulation

RADIC	Radial Distortion Correction
RANSAC	RANdom SAmple Consensus
RMSE	Root Mean Squared Error
ROI	Region of Interest
SAD	Sum of Absolute Differences
SF	Specular-free image
SHD	Sum of Hamming Distances
SIFT	Scale Invariant Feature Transform
SLI	Slit-Lamp Image
SLIM	Slit-Lamp Image Mosaicing
SOM	Self-organizing Map
SPECT	Single-photon Emission Computed Tomography
SS	Self Similarity
SSD	Sum of Squared Differences
SURF	Speeded-Up Robust Features
TC	Tanimoto Coefficient
TPS	Thin-Plate Spline
TRUS	Transrectal Ultrasound
TVUS	Transvaginal Ultrasound
Ugrid	Uniform Grid of points
WHO	World Health Organization

List of Figures

1.1 Important components of the human eye. 2
1.2 Visual signs of diabetic retinopathy and age-related macular degeneration. 3
1.3 Examples of retinal image modalities commonly used for diagnosis and treatment planning. Multimodal imaging performed on a patient's left eye with AMD. (d) shows a zoom-in on a region of IR with a green arrow indicating the corresponding OCT slices of the AMD affected eye (left) and healthy eye (right). 4
1.4 Simplified general illustration of pan-retinal photocoagulation with a contact lens. . . 5
1.5 The main components of the TrackScan platform developed in QuantelMedical and an illustration of SLIM during retinal examination of a patient at University Hospital of Saint-Étienne, France. (a) phantom eye; (b) contact lens; (c) binocular microscope; (d) HD sensors; (e) moving base; (f) laser supply; (g) slit-lamp; (h) live SLI sequence; (k) intra-operative retina map. 7
1.6 TrackScan limitations: uncorrected mosaicing drift. (a) a mismatch of treatment plan (red) with the laser spots (blue); (b) a visually distinctive vessel misalignment. 8
1.7 TrackScan limitations: uncorrected illumination artifacts of different degrees. 8
1.8 TrackScan limitations: multi-modal registration. 9

2.1 General principle of 2D image registration. 15
2.2 Keypoints detected on a photograph of the Cathedral of Clermont Ferrand. SIFT is used to describe the detected keypoints. Yellow circles visualize the position of the keypoint, the scale and orientation of the corresponding descriptor. 17
2.3 Different types of commonly used geometric transformations and their hierarchy. . . 18
2.4 Rigid and parametric non-rigid transformations. 19
2.5 General principle of non-parametric non-rigid transformation. 19
2.6 Different types of the pixel mappings with respect to the similarity metric. 20
2.7 Spatial mapping. 21
2.8 Schematic illustration of the intensity-based image registration approach. 23
2.9 Schematic illustration of the feature-based image registration approach. 23
2.10 Some examples of light-related artifacts in non-medical applications. 27
2.11 Some examples of light-related artifacts in medical imaging applications. 28
2.12 Some examples of light-related artifacts in SLIM. 28

3.1 Retinal image mosaicing results reported by (a) [Pham and Abdollahi, 1991], (b) [Mahurkar et al., 1996] and (c) [Can et al., 2002] respectively. 31
3.2 Retinal image mosaicing results reported by (a) [Stewart et al., 2003], (b) [Yang and Stewart, 2004] and (c) [Choe et al., 2006] respectively. 32

3.3 Retinal image mosaicing results reported by (a) [Cattin et al., 2006], (b) [Aguilar et al., 2007] and (c) [Lee et al., 2008] respectively. 33
3.4 Retinal image mosaicing results reported by (a) [Estrada et al., 2011], (b) [Adal et al., 2014] and (c) [Zheng et al., 2014] respectively. 34
3.5 Retinal image mosaicing results reported by (a) [Asmuth et al., 2001], (b) [Richa et al., 2014] and (c) [Zanet et al., 2016] respectively. 35
3.6 Multi-modal medical image registration results reported by (a) [Tang et al., 2006], (b) [Mitra et al., 2012] and (c) [Yavariabdi et al., 2013] respectively. 40
3.7 Multi-modal retinal image registration results reported by (a) [Matsopoulos et al., 1999], (b) [Choe and Cohen, 2005] and (c) [Broehan et al., 2011] respectively. 41
3.8 Multi-modal retinal image registration results reported by (a) [Chen et al., 2010], (b) [Ghassabi et al., 2013] and (c) [Hernandez et al., 2015] respectively. 42
3.9 Some results on specular highlight correction in medical imaging reported by (a) [Lange, 2005], (b) [Saint-Pierre et al., 2011] and (c) [Allan et al., 2013] respectively. 43
3.10 Some results on specular highlight correction in non-medical applications reported by (a) [Tan and Ikeuchi, 2005], (b) [Yang et al., 2010] and (c [Kim et al., 2013] respectively. 44

4.1 Sample image from each dataset. 52
4.2 Selection of point correspondences. 52
4.3 Error metrics for transformation model complexity evaluation. 55
4.4 The 'spread' evaluation results over different datasets without subsampling. 56
4.5 Examples of registered image pairs with different transformation models. The images are taken from dataset #2. The first image of the sequence is registered with the last image by applying the set of 241 pairwise estimated transformations sequentially. . . 57
4.6 Effect of the number of points. Example of dataset #2. 58

5.1 Examples of mosaics obtained with [Richa et al., 2014]. *(a)* - registration drift is visible through the mismatched vascular structure, *(b)* - example with drift induced blurred regions and duplication. The visual assessment was performed by an expert. 60
5.2 Schematic illustration of track prediction and correction on a sample track τ_j. 62
5.3 Sample images from different slit-lamp datasets. *(a)* - dataset#1, 253 images, *(b)* - dataset#2, 242 images, *(c)* - dataset#3, 169 images, *(d)* - dataset#4, 309 images. . . . 63
5.4 Number of tracks versus frames. Results show performance with UGrid key-points *(red)*, minEig *(green)* and SIFT key-points *(blue)* respectively on the experiment *without* track correction. 65
5.5 Number of tracks versus frames. Results show performance with UGrid key-points *(red)*, minEig *(green)* and SIFT key-points *(blue)* respectively on the experiment *with* track correction. 66
5.6 Examples of improved areas of the mosaics given in Figure 5.1 with corrected drift using proposed approach. First column - originals, second column - corrected versions. 67

6.1 Typical slit-lamp images demonstrating the appearance variation of the light reflection of different origins. CWS - cotton wool spot. 70
6.2 Flowchart of the method to correct light-related imaging artifacts in SLIM. 71
6.3 Schematic illustration of the single-image glare removal and retina segmentation in SLIM. 72
6.4 Schematic illustration of the multi-image lens flare correction in SLIM. 73

6.5 Comparative results of glare removal. GT - Ground Truth. Example of a simple case is shown in the first row and the second row illustrates more complicated condition. Method 1 - [Tan and Ikeuchi, 2005], Method 2 - [Shen et al., 2008], Method 3 - [Yang et al., 2010]. 75
6.6 Comparative results of retinal content segmentation. GT - Ground Truth. Method 1 - thresholding as in SLIM-DF and Method 2 - [Zanet et al., 2016]. 76
6.7 The comparative results for one of the mosaics. Method 1 - SLIM-DF and Method 2 - [Zanet et al., 2016]. 78
6.8 Example of mosaics build with SLIM-DF and corrected with the proposed technique. 79

7.1 Illustration of the inter- and intra-modality geometric and photometric variability between FA (first row) and SLIM (second raw). 82
7.2 Method overview. 83
7.3 The simple approach used to obtain vessel bifurcation points in the FA image. 84
7.4 Template matching scheme for the OD and macula localization. 85
7.5 Example of a result of the retinal features detection in the FA reference image. 85
7.6 An illustration of the training of a SOM. 86
7.7 Schematic overview of the automatic point matching with SOM and LBP. 87
7.8 Sample image pairs from the different datasets. The first column shows the normal retinal image pairs to be registered while the second column illustrates the abnormal cases within a corresponding dataset, indicated row-wise. 91
7.9 Performance of the automatic point matching algorithm between an AF and SLIM sample image pairs from the *FA2SLIM* dataset. Green crosses show the GT point correspondences while blue circles indicate the detected points using the proposed SOM-LBP method. An example with mismatched points is shown in the second row. 93
7.10 Image registration results on sample image pairs from the *FA2SLIM* dataset. From left to right the columns show the reference image with detected landmarks, the target image with detected correspondences (blue) and their GT points (green) and the registration result as a fused image. 96

List of Tables

3.1 Summary of the retinal mosaicing methods. n-D is a multidimensional parameter space. 47
3.2 Summary of the multi-modal retinal registration methods. n-D is a multidimensional parameter space. 48

4.1 Summary of the transformation models' characteristics. The DOF define the number of estimated parameters. We label each model according to whether it is linear w.r.t. its parameters or the source point. k indicates the number of control points of the TPS. 51
4.2 Average ξ_{LFE} and ξ_{LCE} across the different datasets. 55

5.1 Forward-Backward Consistency for similarity metrics evaluation. 64
5.2 Tracking statistics *without* track correction. μ - average number of tracks per frame, κ - number of key-frames, S_{mean} - average span, S_{max} - maximum span. 65
5.3 Tracking statistics *with* track correction. μ - average number of tracks per frame, κ - number of key-frames, S_{mean} - average span, S_{max} - maximum span. 66
5.4 LCE computed across datasets. The proposed (1) is our method with UGrid used for tracks initialization and SHD based local correction step. The proposed (2) is the proposed (1) + local BA. 67

6.1 Retinal content segmentation performance. 74

7.1 The characteristics of the used datasets and corresponding ground truth. $I_{O^{any}}$ denotes the initial number of images where the OD was either visible or not. $I_{O^{vis}}$ denotes the number of images where the OD is fully visible. Subsequent Normal, Abnormal and GT values correspond to $I_{O^{vis}}$. 91
7.2 Evaluation of the effect of the inclusion and exclusion of the macula detection step on the registration errors. 92
7.3 Accumulated statistics over three datasets on Proportion of Correctly matched Keypoints $\mathfrak{M}_{PCK}$ obtained with SOM-LBP. 94
7.4 Mean registration error statistics for the proposed method across three datasets. . . . 95
7.5 Quantitative results of two error metrics, $\mathscr{E}_{RMSE}$ and $\mathscr{E}_{CEM}$, obtained by the proposed method (Angio2SLIM), the semiautomatic method used in TrackScan (Baseline) and the existing algorithms proposed in [Yang et al., 2007] (GDB-ICP), [Markaki et al., 2009] (SOM-09) and [Chen et al., 2010] (PIIFD) for image pairs of the *FA2SLIM* dataset (errors in pixels). 97

Bibliography

[Abraham and Simon, 2013] Abraham, R. and Simon, P. (2013). Review on mosaicing techniques in image processing. In *Advanced Computing and Communication Technologies (ACCT), 2013 Third International Conference on*, pages 63–68. IEEE.

[Adal et al., 2014] Adal, K. M., Ensing, R. M., Couvert, R., Van Etten, P., Martinez, J. P., Vermeer, K. A., and van Vliet, L. J. (2014). A hierarchical coarse-to-fine approach for fundus image registration. In *International Workshop on Biomedical Image Registration*, pages 93–102. Springer.

[Aguilar et al., 2007] Aguilar, W., Martinez-Perez, M. E., Frauel, Y., Escolano, F., Lozano, M. A., and Espinosa-Romero, A. (2007). Graph-based methods for retinal mosaicing and vascular characterization. *Lecture Notes in Computer Science*, 4538:25.

[Ali et al., 2016] Ali, S., Daul, C., Galbrun, E., Guillemin, F., and Blondel, W. (2016). Anisotropic motion estimation on edge preserving riesz wavelets for robust video mosaicing. *Pattern Recognition*, 51:425–442.

[Allan et al., 2013] Allan, M., Ourselin, S., Thompson, S., Hawkes, D. J., Kelly, J., and Stoyanov, D. (2013). Toward detection and localization of instruments in minimally invasive surgery. *IEEE Transactions on Biomedical Engineering*, 60(4):1050–1058.

[Asmuth et al., 2001] Asmuth, J., Madjarov, B., Sajda, P., and Berger, J. W. (2001). Mosaicking and enhancement of slit lamp biomicroscopic fundus images. *British journal of ophthalmology*, 85(5):563–565.

[Bartoli, 2008] Bartoli, A. (2008). Maximizing the predictivity of smooth deformable image warps through cross-validation. *Journal of Mathematical Imaging and Vision*, 31(2-3):133–145.

[Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. *Computer vision–ECCV 2006*, pages 404–417.

[Beaudet, 1978] Beaudet, P. (1978). Rotationally invariant image operators. In *International Joint Conference on Pattern Recognition*, pages 579–583. Kyoto, Japan.

[Besl and McKay, 1992] Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics.

[Björck, 1996] Björck, Å. (1996). *Numerical methods for least squares problems*. SIAM.

[Blumenkranz et al., 2006] Blumenkranz, M. S., Yellachich, D., Andersen, D. E., Wiltberger, M. W., Mordaunt, D., Marcellino, G. R., and Palanker, D. (2006). Semiautomated patterned scanning laser for retinal photocoagulation. *Retina*, 26(3):370–376.

[Boykov and Jolly, 2000] Boykov, Y. and Jolly, M.-P. (2000). Interactive organ segmentation using graph cuts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 276–286. Springer.

[Broehan et al., 2011] Broehan, A. M., Rudolph, T., Amstutz, C. A., and Kowal, J. H. (2011). Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system. *IEEE transactions on biomedical engineering*, 58(10):2816–2824.

[Brown, 1992] Brown, L. G. (1992). A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376.

[Brown and Lowe, 2007] Brown, M. and Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73.

[Calonder et al., 2012] Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., and Fua, P. (2012). Brief: Computing a local binary descriptor very fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1281–1298.

[Can et al., 2002] Can, A., Stewart, C. V., Roysam, B., and Tanenbaum, H. L. (2002). A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):412–419.

[Cattin et al., 2006] Cattin, P. C., Bay, H., Van Gool, L., and Székely, G. (2006). Retina mosaicing using local features. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 185–192. Springer.

[Chakraborty et al., 2014] Chakraborty, B., Marcinczak, J. M., and Grigat, R.-R. (2014). Classification of weak specular reflections in laparoscopic images. In *SPIE Medical Imaging*, pages 90353I–90353I. International Society for Optics and Photonics.

[Chalam et al., 2012] Chalam, K. V., Murthy, R. K., Brar, V., Radhakrishnan, R., Khetpal, V., and Grover, S. (2012). Evaluation of a novel, non contact, automated focal laser with integrated (navilas®) fluorescein angiography for diabetic macular edema. *Middle East African journal of ophthalmology*, 19(1):158.

[Chen et al., 2010] Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R. T., and Laine, A. F. (2010). A partial intensity invariant feature descriptor for multimodal retinal image registration. *IEEE Transactions on Biomedical Engineering*, 57(7):1707–1718.

[Cheng et al., 2016] Cheng, X., Zhang, L., and Zheng, Y. (2016). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pages 1–5.

[Chhablani et al., 2014] Chhablani, J., Mathai, A., Rani, P., Gupta, V., Arevalo, J. F., and Kozak, I. (2014). Comparison of conventional pattern and novel navigated panretinal photocoagulation in proliferative diabetic retinopathycomparison of pascal and navilas for prp. *Investigative ophthalmology & visual science*, 55(6):3432–3438.

[Chhatkuli et al., 2014] Chhatkuli, A., Bartoli, A., Malti, A., and Collins, T. (2014). Live image parsing in uterine laparoscopy. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 1263–1266. IEEE.

[Choe and Cohen, 2005] Choe, T. E. and Cohen, I. (2005). Registration of multimodal fluorescein images sequence of the retina. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 106–113. IEEE.

[Choe et al., 2006] Choe, T. E., Cohen, I., Lee, M., and Medioni, G. (2006). Optimal global mosaic generation from retinal images. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 681–684. IEEE.

[Civera et al., 2009] Civera, J., Davison, A. J., Magallón, J. A., and Montiel, J. (2009). Drift-free real-time sequential mosaicing. *International Journal of Computer Vision*, 81(2):128–137.

[Collins and Bartoli, 2012] Collins, T. and Bartoli, A. (2012). Towards live monocular 3d laparoscopy using shading and specularity information. *Information Processing in Computer-Assisted Interventions*, pages 11–21.

[Collins et al., 2014] Collins, T., Pizarro, D., Bartoli, A., Canis, M., and Bourdel, N. (2014). Computer-assisted laparoscopic myomectomy by augmenting the uterus with pre-operative mri data. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 243–248. IEEE.

[Davis, 1998] Davis, J. (1998). Mosaics of scenes with moving objects. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 354–360. IEEE.

[DRS, 1981] DRS (1981). Photocoagulation treatment of proliferative diabetic retinopathy: clinical application of diabetic retinopathy study (drs) findings, drs report number 8. *Ophthalmology*, 88(7):583–600.

[Estrada et al., 2011] Estrada, R., Tomasi, C., Cabrera, M. T., Wallace, D. K., Freedman, S. F., and Farsiu, S. (2011). Enhanced video indirect ophthalmoscopy (vio) via robust mosaicing. *Biomedical optics express*, 2(10):2871–2887.

[ETDRS, 1991] ETDRS (1991). Early photocoagulation for diabetic retinopathy: Etdrs report number 9. *Ophthalmology*, 98(5):766–785.

[Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

[Ghassabi et al., 2013] Ghassabi, Z., Shanbehzadeh, J., Sedaghat, A., and Fatemizadeh, E. (2013). An efficient approach for robust multimodal retinal image registration based on ur-sift features and piifd descriptors. *EURASIP Journal on Image and Video Processing*, 2013(1):1–16.

[Ghosh and Kaabouch, 2016] Ghosh, D. and Kaabouch, N. (2016). A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34:1–11.

[Goldberg, 1989] Goldberg, D. (1989). Genetic algorithms in optimization, search and machine learning. *Reading: Addison-Wesley*.

[Gutiérrez-Becker et al., 2016] Gutiérrez-Becker, B., Mateus, D., Peter, L., and Navab, N. (2016). Learning optimization updates for multimodal registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 19–27. Springer.

[Harris and Stephens, 1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Manchester, UK.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

[Hartley, 1997] Hartley, R. I. (1997). Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1):5–23.

[He et al., 2012] He, Y., Khanna, N., Boushey, C. J., and Delp, E. J. (2012). Specular highlight removal for image-based dietary assessment. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, pages 424–428. IEEE.

[Hellier and Barillot, 2004] Hellier, P. and Barillot, C. (2004). A hierarchical parametric algorithm for deformable multimodal image registration. *Computer Methods and Programs in Biomedicine*, 75(2):107–115.

[Hernandez et al., 2015] Hernandez, M., Medioni, G., Hu, Z., and Sadda, S. (2015). Multimodal registration of multiple retinal images based on line structures. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pages 907–914. IEEE.

[Hernandez-Matas et al., 2016] Hernandez-Matas, C., Zabulis, X., and Argyros, A. A. (2016). Retinal image registration through simultaneous camera pose and eye shape estimation. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 3247–3251. IEEE.

[Ho, 1995] Ho, T. K. (1995). Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE.

[Horn and Schunck, 1981] Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, 17(1-3):185–203.

[Howe, 2015] Howe, N. R. (2015). Contour-pruned skeletonization. http://www.cs.smith.edu/~nhowe/research/code/.

[Inan et al., 2016] Inan, U. U., Polat, O., Inan, S., Yigit, S., and Baysal, Z. (2016). Comparison of pain scores between patients undergoing panretinal photocoagulation using navigated or pattern scan laser systems. *Arquivos brasileiros de oftalmologia*, 79(1):15–18.

[Ingber and Rosen, 1992] Ingber, L. and Rosen, B. (1992). Genetic algorithms and very fast simulated reannealing: A comparison. *Mathematical and computer modelling*, 16(11):87–100.

[Irani et al., 1995] Irani, M., Anandan, P., and Hsu, S. (1995). Mosaic based representations of video sequences and their applications. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 605–611. IEEE.

[Kadir and Brady, 2001] Kadir, T. and Brady, M. (2001). Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105.

[Khaderi et al., 2011] Khaderi, K. R., Ahmed, K. A., Berry, J. L., Labriola, L. T., and Cornwell, R. (2011). Retinal imaging modalities: advantages and limitations for clinical practice. *Retinal Physician*, 8(3):44–46.

[Kim et al., 2013] Kim, H., Jin, H., Hadap, S., and Kweon, I. (2013). Specular reflection separation using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1460–1467.

[Kingslake, 1992] Kingslake, R. (1992). *Optics in photography*, volume 6. SPIE Press.

[Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802.

[Klein and Murray, 2007] Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small AR workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234.

[Köhler et al., 2016] Köhler, T., Heinrich, A., Maier, A., Hornegger, J., and Tornow, R. P. (2016). Super-resolved retinal image mosaicing. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1063–1067. IEEE.

[Kohonen, 1998] Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21(1):1–6.

[Konolige and Agrawal, 2008] Konolige, K. and Agrawal, M. (2008). FrameSLAM: From bundle adjustment to real-time visual mapping. *Robotics, IEEE Transactions on*, 24(5):1066–1077.

[Kozak and Luttrull, 2015] Kozak, I. and Luttrull, J. K. (2015). Modern retinal laser therapy. *Saudi Journal of Ophthalmology*, 29(2):137 – 146.

[Kozak et al., 2011] Kozak, I., Oster, S. F., Cortes, M. A., Dowell, D., Hartmann, K., Kim, J. S., and Freeman, W. R. (2011). Clinical evaluation and treatment accuracy in diabetic macular edema using navigated laser photocoagulator navilas. *Ophthalmology*, 118(6):1119 – 1124.

[Kumar et al., 1995] Kumar, R., Anandan, P., Irani, M., Bergen, J., and Hanna, K. (1995). Representation of scenes from collections of images. In *Representation of Visual Scenes, 1995.(In Conjuction with ICCV'95), Proceedings IEEE Workshop on*, pages 10–17. IEEE.

[Laliberté et al., 2003] Laliberté, F., Gagnon, L., and Sheng, Y. (2003). Registration and fusion of retinal images-an evaluation study. *IEEE Transactions on Medical Imaging*, 22(5):661–673.

[Lange, 2005] Lange, H. (2005). Automatic glare removal in reflectance imagery of the uterine cervix. In *Medical Imaging*, pages 2183–2192. International Society for Optics and Photonics.

[Lawson and Hanson, 1974] Lawson and Hanson (1974). *Solving least squares problems*, volume 161. Prentice-hall.

[Lee et al., 2007] Lee, S., Abràmoff, M. D., and Reinhardt, J. M. (2007). Validation of retinal image registration algorithms by a projective imaging distortion model. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 6471–6474. IEEE.

[Lee et al., 2008] Lee, S., Abràmoff, M. D., and Reinhardt, J. M. (2008). Retinal image mosaicing using the radial distortion correction model. In *Proc. of SPIE Vol*, volume 6914, pages 691435–1.

[Li et al., 2008a] Li, J., Chen, H., Yao, C., and Zhang, X. (2008a). A robust feature-based method for mosaic of the curved human color retinal images. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, volume 1, pages 845–849. IEEE.

[Li et al., 2008b] Li, Y., Wang, Y., Huang, W., and Zhang, Z. (2008b). Automatic image stitching using sift. In *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on*, pages 568–571. IEEE.

[Linhares et al., 2016] Linhares, R., Richa, R., de Moraes, R., Sobieranski, A., and von Wangenheim, A. (2016). Non-rigid fine adjustment of retina maps acquired using a slit-lamp. In *Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on*, pages 285–289. IEEE.

[Livyatan et al., 2003] Livyatan, H., Yaniv, Z., and Joskowicz, L. (2003). Gradient-based 2-d/3-d rigid registration of fluoroscopic x-ray to ct. *IEEE Transactions on Medical Imaging*, 22(11):1395–1406.

[Lovegrove and Davison, 2010] Lovegrove, S. and Davison, A. J. (2010). Real-time spherical mosaicing using whole image alignment. In *European Conference on Computer Vision*, pages 73–86. Springer.

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

[Lu and Milios, 1997] Lu, F. and Milios, E. (1997). Globally consistent range scan alignment for environment mapping. *Autonomous robots*, 4(4):333–349.

[Maes et al., 1997] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2):187–198.

[Mahurkar et al., 1996] Mahurkar, A. A., Vivino, M. A., Trus, B. L., Kuehl, E. M., Datiles, M., and Kaiser-Kupfer, M. I. (1996). Constructing retinal fundus photomontages. a new computer-based method. *Investigative Ophthalmology & Visual Science*, 37(8):1675–1683.

[Maintz and Viergever, 1998] Maintz, J. A. and Viergever, M. A. (1998). A survey of medical image registration. *Medical image analysis*, 2(1):1–36.

[Mani and Rivazhagan, 2013] Mani, V. and Rivazhagan, D. (2013). Survey of medical image registration. *Journal of Biomedical Engineering and Technology*, 1(2):8–25.

[Maninis et al., 2016] Maninis, K., Pont-Tuset, J., Arbeláez, P., and Gool, L. V. (2016). Deep retinal image understanding. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.

[Mann and Picard, 1995] Mann, S. and Picard, R. W. (1995). *Video orbits of the projective group: A new perspective on image mosaicing*. Perceptual Computing Section, Media Laboratory, Massachusetts Institute of Technology.

[Markaki et al., 2009] Markaki, V. E., Asvestas, P. A., and Matsopoulos, G. K. (2009). Application of kohonen network for automatic point correspondence in 2d medical images. *Computers in Biology and Medicine*, 39(7):630–645.

[Marzotto et al., 2004] Marzotto, R., Fusiello, A., and Murino, V. (2004). High resolution video mosaicing with global alignment. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE.

[Matas et al., 2004] Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10):761–767.

[Matsopoulos et al., 2004] Matsopoulos, G. K., Asvestas, P. A., Mouravliansky, N. A., and Delibasis, K. K. (2004). Multimodal registration of retinal images using self organizing maps. *IEEE Transactions on Medical Imaging*, 23(12):1557–1563.

[Matsopoulos et al., 1999] Matsopoulos, G. K., Mouravliansky, N. A., Delibasis, K. K., and Nikita, K. S. (1999). Automatic retinal image registration scheme using global optimization techniques. *IEEE Transactions on Information Technology in Biomedicine*, 3(1):47–60.

[McLauchlan and Jaenicke, 2002] McLauchlan, P. F. and Jaenicke, A. (2002). Image mosaicing using sequential bundle adjustment. *Image and Vision computing*, 20(9):751–759.

[Mikolajczyk and Schmid, 2004] Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1):63–86.

[Mitra et al., 2012] Mitra, J., Martí, R., Oliver, A., Lladó, X., Ghose, S., Vilanova, J. C., and Meriaudeau, F. (2012). Prostate multimodality image registration based on b-splines and quadrature local energy. *International journal of computer assisted radiology and surgery*, 7(3):445–454.

[Moré, 1978] Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

[Mouragnon et al., 2009] Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., and Sayd, P. (2009). Generic and real-time structure from motion using local bundle adjustment. *Image and Vision Computing*, 27(8):1178–1193.

[Muja and Lowe, 2009] Muja, M. and Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP (1)*, 2(331-340):2.

[Müller et al., 2011] Müller, K., Bauer, S., Wasza, J., and Hornegger, J. (2011). Automatic multimodal tof/ct organ surface registration. *Bildverarbeitung für die Medizin 2011*, pages 154–158.

[Nayar et al., 1993] Nayar, S. K., Fang, X.-S., and Boult, T. (1993). Removal of specularities using color and polarization. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 583–590. IEEE.

[Nister and Stewenius, 2006] Nister, D. and Stewenius, H. (2006). Scalable recognition with a vocabulary tree. In *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, volume 2, pages 2161–2168. Ieee.

[Nussberger et al., 2015] Nussberger, A., Grabner, H., and Van Gool, L. (2015). Robust aerial object tracking in images with lens flare. In *2015 IEEE International Conference on Robotics and Automation*, pages 6380–6387. IEEE.

[Ojala et al., 2002] Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987.

[Oktay et al., 2015] Oktay, O., Schuh, A., Rajchl, M., Keraudren, K., Gomez, A., Heinrich, M. P., Penney, G., and Rueckert, D. (2015). Structured decision forests for multi-modal ultrasound image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 363–371. Springer.

[Osadchy et al., 2003] Osadchy, M., Jacobs, D., and Ramamoorthi, R. (2003). Using specularities for recognition. In *null*, page 1512. IEEE.

[Pascolini and Mariotti, 2011] Pascolini, D. and Mariotti, S. P. (2011). Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, pages bjophthalmol–2011.

[Pham and Abdollahi, 1991] Pham, D. T. and Abdollahi, M. (1991). Automatic assembly of ocular fundus images. *Pattern recognition*, 24(3):253–262.

[Pluim et al., 2003] Pluim, J. P., Maintz, J. A., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004.

[Powell, 1964] Powell, M. J. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162.

[Prokopetc and Bartoli, 2016a] Prokopetc, K. and Bartoli, A. (2016a). A comparative study of transformation models for the sequential mosaicing of long retinal sequences of slit-lamp images obtained in a closed-loop motion. *International journal of computer assisted radiology and surgery*, 11(12):2163–2172.

[Prokopetc and Bartoli, 2016b] Prokopetc, K. and Bartoli, A. (2016b). Reducing drift in mosaicing slit-lamp retinal images. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on*, pages 533–540. IEEE.

[Prokopetc and Bartoli, 2017a] Prokopetc, K. and Bartoli, A. (2017a). Slim: slit lamp image mosaicing. In *Congrés Francophone des Jeunes Chercheurs en Vision par Ordinateur (ORASIS'17)*.

[Prokopetc and Bartoli, 2017b] Prokopetc, K. and Bartoli, A. (2017b). Slim (slit lamp image mosaicing): handling reflection artifacts. *International journal of computer assisted radiology and surgery*, 12(6):911–920.

[Reinhard et al., 2002] Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276.

[Richa et al., 2014] Richa, R., Linhares, R., Comunello, E., Von Wangenheim, A., Schnitzler, J.-Y., Wassmer, B., Guillemot, C., Thuret, G., Gain, P., Hager, G., et al. (2014). Fundus Image Mosaicking for Information Augmentation in Computer-Assisted Slit-Lamp Imaging. *Medical Imaging, IEEE Transactions on*, 33(6):1304–1312.

[Richa et al., 2012] Richa, R., Vágvölgyi, B., Balicki, M., Hager, G., and Taylor, R. H. (2012). Hybrid tracking and mosaicking for information augmentation in retinal surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 397–404. Springer.

[Ridler and Calvard, 1978] Ridler, T. and Calvard, S. (1978). Picture thresholding using an iterative selection method. *IEEE transactions on Systems, Man and Cybernetics*, 8(8):630–632.

[Roche et al., 2001] Roche, A., Pennec, X., Malandain, G., and Ayache, N. (2001). Rigid registration of 3-d ultrasound with mr images: a new approach combining intensity and gradient information. *IEEE transactions on medical imaging*, 20(10):1038–1049.

[Rohlfing et al., 2003] Rohlfing, T., Maurer, C. R., Bluemke, D. A., and Jacobs, M. A. (2003). Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. *IEEE transactions on medical imaging*, 22(6):730–741.

[Rueckert et al., 1999] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721.

[Saint-Pierre et al., 2011] Saint-Pierre, C.-A., Boisvert, J., Grimard, G., and Cheriet, F. (2011). Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Machine Vision and Applications*, 22(1):171–180.

[Sawhney et al., 1998] Sawhney, H. S., Hsu, S., and Kumar, R. (1998). Robust video mosaicing through topology inference and local to global alignment. In *European conference on computer vision*, pages 103–119. Springer.

[Sawhney and Kumar, 1999] Sawhney, H. S. and Kumar, R. (1999). True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):235–243.

[Shen et al., 2008] Shen, H.-L., Zhang, H.-G., Shao, S.-J., and Xin, J. H. (2008). Chromaticity-based separation of reflection components in a single image. *Pattern Recognition*, 41(8):2461–2469.

[Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600.

[Shirin et al., 2012] Shirin, H. M. A., Rabbani, H., and Akhlaghi, M. R. (2012). Diabetic retinopathy grading by digital curvelet transform. *Computational and mathematical methods in medicine*, 2012.

[Simonovsky et al., 2016] Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. (2016). A deep metric for multimodal registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 10–18. Springer.

[Slama et al., 1980] Slama, C. C., Theurer, C., and Henriksen, S. W. (1980). *Manual of photogrammetry*. American Society of photogrammetry.

[Souza et al., 2014] Souza, M., Richa, R., Puel, A., Caetano, J., Comunello, E., and Von Wangenheim, A. (2014). Robust visual tracking for retinal mapping in computer-assisted slit-lamp imaging. In *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, pages 132–137. IEEE.

[Steedly et al., 2005] Steedly, D., Pal, C., and Szeliski, R. (2005). Efficiently registering video into panoramic mosaics. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1300–1307. IEEE.

[Stewart et al., 2003] Stewart, C. V., Tsai, C.-L., and Roysam, B. (2003). The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE transactions on medical imaging*, 22(11):1379–1394.

[Stewart, 2017] Stewart, M. W. (2017). *Diabetic Retinopathy: Current Pharmacologic Treatment and Emerging Strategies*. Springer.

[Stoyanov et al., 2005] Stoyanov, D., Darzi, A., and Yang, G. Z. (2005). A practical approach towards accurate dense 3d depth recovery for robotic laparoscopic surgery. *Computer Aided Surgery*, 10(4):199–208.

[Sun et al., 2016] Sun, C., Liu, S., Yang, T., Zeng, B., Wang, Z., and Liu, G. (2016). Automatic reflection removal using gradient intensity and motion cues. In *Proceedings of the ACM on Multimedia Conference*, pages 466–470. ACM.

[Szeliski, 2006] Szeliski, R. (2006). Image alignment and stitching: A tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104.

[Szeliski and Shum, 1997] Szeliski, R. and Shum, H.-Y. (1997). Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 251–258. ACM Press/Addison-Wesley Publishing Co.

[Tan and Ikeuchi, 2005] Tan, R. T. and Ikeuchi, K. (2005). Separating reflection components of textured surfaces using a single image. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 27(2):178–193.

[Tanaka et al., 1978] Tanaka, M., Tamura, S., and Tanaka, K. (1978). A technique for the automatic assembly of eye fundus photographs using blood vessel structure. In *Proceedings of the International Conference on Cybernetics and Society, Tokyo and Kyoto*, pages 266–270.

[Tang et al., 2006] Tang, L., Hamarneh, G., and Celler, A. (2006). Co-registration of bone ct and spect images using mutual information. In *Signal Processing and Information Technology, 2006 IEEE International Symposium on*, pages 116–121. IEEE.

[Tomasi and Kanade, 1991] Tomasi, C. and Kanade, T. (1991). *Detection and tracking of point features*.

[Triggs et al., 1999] Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (1999). Bundle adjustment—a modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372.

[Tuytelaars and Van Gool, 2004] Tuytelaars, T. and Van Gool, L. (2004). Matching widely separated views based on affine invariant regions. *International journal of computer vision*, 59(1):61–85.

[Van Nguyen et al., 2015] Van Nguyen, H., Zhou, K., and Vemulapalli, R. (2015). Cross-domain synthesis of medical images using efficient location-sensitive deep network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 677–684. Springer.

[Vialard et al., 2012] Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J. (2012). Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241.

[Viergever et al., 2016] Viergever, M. A., Maintz, J. A., Klein, S., Murphy, K., Staring, M., and Pluim, J. P. (2016). A survey of medical image registration–under review. *Medical Image Analysis*, 33:140–144.

[Wang et al., 2004] Wang, J., Eng, H.-L., Kam, A. H., and Yau, W.-Y. (2004). Specular reflection removal for human detection under aquatic environment. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pages 130–130. IEEE.

[Wang et al., 2010] Wang, Y., Shen, J., Liao, W., and Zhou, L. (2010). Automatic fundus images mosaic based on sift feature. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 6, pages 2747–2751. IEEE.

[Wei et al., 2009] Wei, L., Huang, L., Pan, L., and Yu, L. (2009). The retinal image mosaic based on invariant feature and hierarchial transformation models. In *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*, pages 1–5. IEEE.

[Wright et al., 2000] Wright, C. H., Barrett, S. F., Ferguson, R. D., Rylander III, H. G., and Welch, A. J. (2000). Initial in vivo results of a hybrid retinal photocoagulation system. *Journal of biomedical optics*, 5(1):56–61.

[Xu et al., 2015] Xu, C., Wang, X., Wang, H., and Zhang, Y. (2015). Accurate image specular highlight removal based on light field imaging. In *Visual Communications and Image Processing*, pages 1–4. IEEE.

[Xu, 2013] Xu, Z. (2013). Consistent image alignment for video mosaicing. *Signal, Image and Video Processing*, pages 1–7.

[Yang and Stewart, 2004] Yang, G. and Stewart, C. V. (2004). Covariance-driven mosaic formation from sparsely-overlapping image sets with application to retinal image mosaicing. In *Computer Vision and Pattern Recognition (CVPR), Proceedings of the IEEE Computer Society Conference on*, volume 1, pages I–804. IEEE.

[Yang et al., 2007] Yang, G., Stewart, C. V., Sofka, M., and Tsai, C.-L. (2007). Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE transactions on pattern analysis and machine intelligence*, 29(11).

[Yang et al., 2010] Yang, Q., Wang, S., and Ahuja, N. (2010). Real-time specular highlight removal using bilateral filtering. In *European Conference on Computer Vision*, pages 87–100. Springer.

[Yang et al., 2011] Yang, Q., Wang, S., Ahuja, N., and Yang, R. (2011). A uniform framework for estimating illumination chromaticity, correspondence, and specular reflection. *IEEE Transactions on Image Processing*, 20(1):53–63.

[Yang et al., 2017] Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Fast predictive multimodal image registration. *arXiv preprint arXiv:1703.10902*.

[Yao, 2008] Yao, L. (2008). Image mosaic based on sift and deformation propagation. In *Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. IEEE International Symposium on*, pages 848–851. IEEE.

[Yavariabdi et al., 2013] Yavariabdi, A., Samir, C., Bartoli, A., Da Ines, D., and Bourdel, N. (2013). Contour-based tvus-mr image registration for mapping small endometrial implants. In *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, pages 145–154. Springer.

[Zanet et al., 2016] Zanet, S., Rudolph, T., Richa, R., Tappeiner, C., and Sznitman, R. (2016). Retinal slit lamp video mosaicking. *Int. J. of Computer Assisted Radiology and Surgery*, pages 1–7.

[Zelnik-Manor and Irani, 2000] Zelnik-Manor, L. and Irani, M. (2000). Multi-frame estimation of planar motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1105–1116.

[Zheng et al., 2014] Zheng, Y., Daniel, E., Hunter, A. A., Xiao, R., Gao, J., Li, H., Maguire, M. G., Brainard, D. H., and Gee, J. C. (2014). Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix. *Medical image analysis*, 18(6):903–913.

[Zitova and Flusser, 2003] Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.

[Zuiderveld, 1994] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc.