FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Mobile-based Risk Assessment of Diabetic Retinopathy by Image Processing

João Miguel Ribeiro da Costa

Master in Bioengineering

Supervisor: António Miguel Pontes Pimenta Monteiro (PhD)

Supervisor at Fraunhofer: Filipe Cruz Gomes Soares (PhD)

July 2015

Mobile-based Risk Assessment of Diabetic Retinopathy by Image Processing

João Miguel Ribeiro da Costa

Master in Bioengineering

Resumo

A retinopatia diabética é uma condição ocular relacionada com aspetos sistémicos em pacientes diabéticos (hipertensão, controlo glicémico deficiente), em que a retina é progressivamente obstruída por depósitos lipídicos, edemas, etc. Caso não seja tratada de forma adequada, esta condição pode levar à cegueira total.

A progressão inicial é assimtomática, e alterações na função visual geralmente não são detetadas antes de estágios avançados, onde reverter essas alterações frequentemente já não é possível. Contudo, o tratamento precoce tem-se mostrado eficaz em parar a progressão da doença.

Devido à elevada incidência de retinopatia diabética em pacientes diabéticos (para os que sofrem de diabetes há mais de 20 anos, a taxa de incidência é de 76%), é recomendado um rastreio anual através de um exame oftalmoscópico. Contudo, o elevado número de pacientes diabéticos fazem com que esta recomendação seja díficil de cumprir.

Neste projeto procurou-se desenvolver uma solução oftalmoscópica móvel capaz de efetuar deteção automática de estágios iniciais de retinopatia diabética, especificamente microaneurismas. O objetivo desta solução é permitir ao paciente diabético e/ou a médicos de clínica geral efetuarem a aquisição de imagens oftalmoscópicas.

A solução desenvolvida tem como alvo os smartphones Android (equipados com um adaptador oftalmoscópico) e, através da localização do disco óptico por técnicas de visão computacional, auxiliar o utilizador na aquisição das imagens.

Foi elaborado um protocolo de aquisição de forma a assegurar que as imagens que irão ser analisadas são adquiridas em condições semelhantes, garantindo assim o desempenho adequado do algoritmo de deteção. Um método para fundir várias imagens foi também desenvolvido, de forma a aumentar a área da retina que pode ser analisada.

O dataset Retinopathy Online Challenge foi utlizado para efetuar a validação do algoritmo de deteção de microaneurismas. Os resultados evidenciam que o método elaborado tem desempenho comparável a outros algoritmos do estado da arte, tanto no que diz respeito aos resultados finais da deteção, como na comparação entre estágios individuais dos métodos.

Abstract

Diabetic retinopathy is a condition of the eye, related with systemic aspects often verified in diabetes mellitus patients (hypertension, poor glycemic control), where the retina is progressively occluded by lipid deposits, oedemas, etc. Untreated, this condition may ultimately lead to complete vision loss.

Initial progression is asymptomatic, and changes in the visual function are generally not detected before advanced stages, where reversing those changes is often impossible. However, early treatment has been proven effective in halting disease progression.

Due to the large incidence of diabetic retinopathy among diabetic patients (for those with diabetes mellitus for more than 20 years, the incidence rate is 76%), yearly screening, through an ophthalmoscopic examination, is recommended. However, the large number of diabetic patients make this recommendation difficult to follow.

In this project, a mobile ophthalmoscopic solution capable of performing automatic detection of early diabetic retinopathy signs (retinal microaneurysms) was developed. The purpose of this solution is to allow assessment of diabetic retinopathy condition by the diabetic patient and/or by primary care physicians.

The developed solution targets Android smartphones (equipped with an ophthalmoscopic adapter) and, through the localization of the optic disc using computer vision techniques, also aids the user in the acquisition of the retinal images.

An acquisition protocol was devised, in order to ensure that the images to be analysed are acquired in similar conditions. A stitching method for retinal images was also developed, so that the available FOV of the retina in each image is expanded, by merging complementary images.

The Retinopathy Online Challenge dataset was chosen to assess the performance of the developed microaneurysm detection algorithm. The results evidentiate that the devised method performs comparably to state of the art methods, either in respect to the overall detection results or when comparing individual stages of the detection pipeline.

"Everything yields to diligence."

Antiphanes

Contents

List of Figures

List of Tables

Abbreviations and Symbols

CAD	Computer Aided Diagnosis
CHT	Circle Hough Transform
CRA	Central Retinal Artery
DR	Diabetic Retinopathy
ER	Entity Relationship
DRIVE	Digital Retinal Images for Vessel Extraction
FOV	Field Of View
FROC	Free Response Receiver Operating Characteristic
FP	False Positive
FN	False Negative
GC	Gaussian Correlation
IR	Infrared
JNI	Java Native Interface
MA	Microaneurysm
MLP	Multilayer Perceptron
NCM	Non Corrected Morphological Vessel Extraction
NPDR	Non Proliferative Diabetic Retinopathy
ORB	Oriented BRIEF
PDR	Proliferative Diabetic Retinopathy
RANSAC	RANdom SAmple Consensus
RBF	Radial Basis Function
ROC	Retinopathy Online Challenge
SE	Structuring Element
SCM	Shade Corrected Morphological Vessel Extraction
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
TP	True Positive
TN	True Negative

Chapter 1

Introduction

In this work, a mobile solution capable of detecting early signs of diabetic retinopathy (DR) is developed. This is achieved based on the detection of microaneurysms in images of the eye fundus, obtained by mobile ophthalmoscopy. Microaneurysms are one of the earliest symptoms of DR and significant literature supports microaneurysms count as a robust indicator of the severity of DR [2, 3].

A particular important aspect of DR is the limited available options for treatment of advanced DR stages, which leads to difficulty in reverting vision threatening changes. Early diagnosis is thus very important in order to avoid serious vision loss, as detailed in section 2.2.3. However, the problem of early diagnosis is aggravated by sparse examinations, and the difficulty in implementing a large scale screening program. This project addresses this problem, by allowing the patients to perform regular self-assessment of their condition.

The mobile solution is implemented in Android smartphones and in order to allow image acquisition of the eye fundus, an ophthalmoscopic adapter to the smartphone camera is used. A proper review of the available ophthalmoscopic adapters is presented in section 3.1.

1.1 Context and Motivation

It is estimated that in 2010, there were 285 million people worldwide diagnosed with diabetes and, from those, one third presented signs of diabetic retinopathy [4]. Furthermore, the increasing prevalence of diabetes in the global population as well as the increase in life expectancy of the patients are causing diabetic retinopathy to be the leading cause of vision loss in several countries [5].

Prevalence of DR in the diabetic population is similar across different countries (around a third of the total diabetic population). Developed and developing countries have shown similar rates, but some Asian regions present higher rates, above 40%. In relation to age,

data for the US has shown similar rates for diabetic patients with age between 40 and 64 (28%) and 65+ (30%) [5]. A major differentiator in the incidence of DR is the duration of diabetes, as patients with diabetes for less than ten years have a prevalence around 21% and those with diabetes for more than twenty years present a rate of 76% [5]. For this reason, this is one of the factors weighted in the design of screening programmes, as explained in section 2.2.5.

The epidemiological characteristics of DR serve as motivation for a number of approaches that try to improve the screening process, using automated methods for the detection in ophthalmoscopic/angiographic images. These methods are rooted in traditional computer vision and machine learning approaches.

This project will also be inserted in the computer vision and machine learning fields, but it will try to tackle a larger problem than the recognition of early indicators of the disease: besides this task, the solution should comprise the proper acquisition and management of ophthalmoscopic images, through an Android application. In fact, by exploring the significant computational power and image acquisition capabilities of modern day smartphones, it is possible to have this complete solution in a mobile context, with relatively low costs for the user - only a smartphone and an ophthalmoscopic adapter are needed.

The project is motivated not only by the large diabetic population (which, as the presented statistics show, have a high propensity for developing DR), but also for the significant unresolved challenges in the screening process, including the lack of available ophthalmologists and difficulties in meeting the recommended yearly periodic screening of each diabetic patient. These shortcomings may significantly increase the risk of DR in the diabetic population, particularly if some risk factors are not properly accounted for. A low cost, reliable system that could perform image acquisition of the eye fundus and also a preliminary detection of DR signs could help prevent serious vision loss, as the changes caused by DR are often irreversible.

1.2 Objectives

In the development of this project, four objectives will be targeted:

- Definition of an image acquisition protocol for smartphone fundography.
- Computer aided alignment of the eye during smartphone image acquisition.
- Increase available retinal FOV through image stitching.
- Image recognition and automatic annotation of microaneurysms on eye fundus.

A proper definition of an acquisition protocol for smartphone fundography is essential for the success of this project. In fact, the small size of microaneurysms, besides rendering

their detection very susceptible to noise, also require high magnification and resolution in the image acquisition. Even though these factors will be properly weighted in the choice of the ophthalmoscopic adapter, performance at par with fundus photography cameras will be difficult to match, and thus an appropriate acquisition protocol may help mitigate these issues.

Alignment of the eye fundus in the acquisition should be enforced due to the fact that clinical significant microaneurysms are mostly located around specific retinal regions (as will be explained in section 2.2). Alignment of the eye fundus can be performed by detecting this structure (macula) or, alternatively, the optic disc. The problem of detecting this structure is completely different from the main problem that is the detection of microaneurysms, even though it will also be approached by a computer vision perspective. This task should be performed in real time, in order to allow automatic guidance to the examiner.

The constrained retinal FOV available through the use of smartphone adapters poses a problem for the examination. As such, image stitching should be used to increase this FOV, by merging images of different regions of the retina.

The recognition of microaneurysms in the acquired image is the main objective of this work. The implementation of the detection method will be achieved using the *OpenCV* framework and this implementation should be adapted to the limited computational power available in smartphones.

1.3 Overview

This work is divided in six chapters. Following this initial chapter, which introduces the theme, chapter 2 presents a literature review of several aspects related to DR as well as to related work in this field.

In chapter 3, the methodology and tools used in this work are overviewed. The datasets used for validation are described, the different smartphone ophthalmoscopic adapters are compared, image stitching techniques are introduced and a brief review of classification is presented.

In chapter 4, an exam guidance procedure is described, along with a method to locate the optic disc and another method to detect the FOV of the ophthalmoscopic adapter. The information of the location of the optic disc and the FOV of the adapter are used to guide the user in the acquisition.

In chapter 5, the retinal stitching procedure is analysed. Different approaches for the computation of the alignment are compared, along with the choice of vessel enhancement methods that most improve the alignment process.

In chapter 6, the microaneurysm detection pipeline, along with its various processing stages and respective results are described. These stages include the extraction of candidates from the original images, the segmentation of retinal vasculature to eliminate false candidates and the classification of those candidates as true/false MAs, after the computation of candidate features.

In chapter 7, the developed Android application is reviewed, with focus on its structure and implementation details.

Finally, in chapter 8 the conclusion of this work is presented along with some topics regarding future developments.

Chapter 2

Diabetic Retinopathy Characterization and Detection

2.1 Anatomy of the Eye

In order to introduce the theme and provide a basic understanding of the human visual system, a brief introduction to the anatomy of the eye will be made.

The anatomy of the eye is conventionally divided in three layers (see Figure 2.1) [6]:

1. Fibrous layer
2. Vascular layer
3. Nervous layer

2.1.1 Fibrous Layer

The fibrous layer is the outermost coat of the eye and has a protective function. It includes two structures: the sclera and the cornea. The sclera is a white, opaque and hard structure that encloses most of the eye. It is responsible not only for the shape of the eye, but also for providing an anchor for the muscles responsible for the eye movements (saccades).

The cornea is also part of the outermost coat of the eye and is a transparent membrane that covers important eye structures such as the iris or the pupil. Despite its protective function, the cornea also has an important optic function, being responsible for most of the eye optic power.

2.1.2 Vascular Layer

The vascular layer includes (from anterior to posterior location) the iris, the ciliary body and the choroid. The iris is a circular, pigmented structure, responsible for controlling the amount of light received in the retina, through the adjustment of the diameter of the pupil.

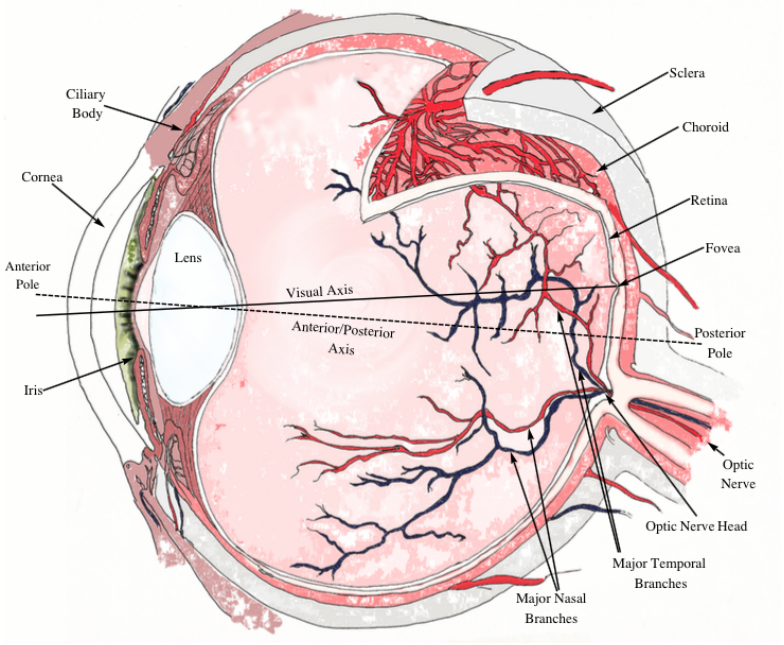


Figure 2.1: Schematic image of eye structures. Image from [6].

The ciliary body is a structure with diverse functions, the most important being the accommodation of the eye. Accommodation consists in maintaining a clear image of an object at different distances. This is possible, through an associated muscle, the ciliary muscle, which contracts or relaxes, making the lens more or less convex, thus changing the total optical power of the eye.

The choroid is located between the sclera and the retina, and is responsible for the nourishment of the outer retina.

2.1.3 Nervous Layer

The nervous layer includes the retina, which is a light sensitive tissue, responsible for transforming the incident light into nervous stimulus, in order to be processed by the brain. The retina is located in the posterior region of the eyeball and is constituted by photosensitive cells: rods and cones. In a general way, rods are responsible for low light and peripheral vision, being located in the outer edges of the retina, while cones are responsible for high resolution and color vision. Cones are concentrated in a specific region of the retina, the fovea, which is an avascular region in the center of the macula. Due to the high concentration of photosensitive cells in this region, the fovea is responsible for the sharp central vision.

The enervation and vascularization of the retina is performed mostly through a region known as the optic disk, which appears as a bright circular spot in fundus images (see Figure 2.2). This is a region where there are no photosensitive cells, and due to this

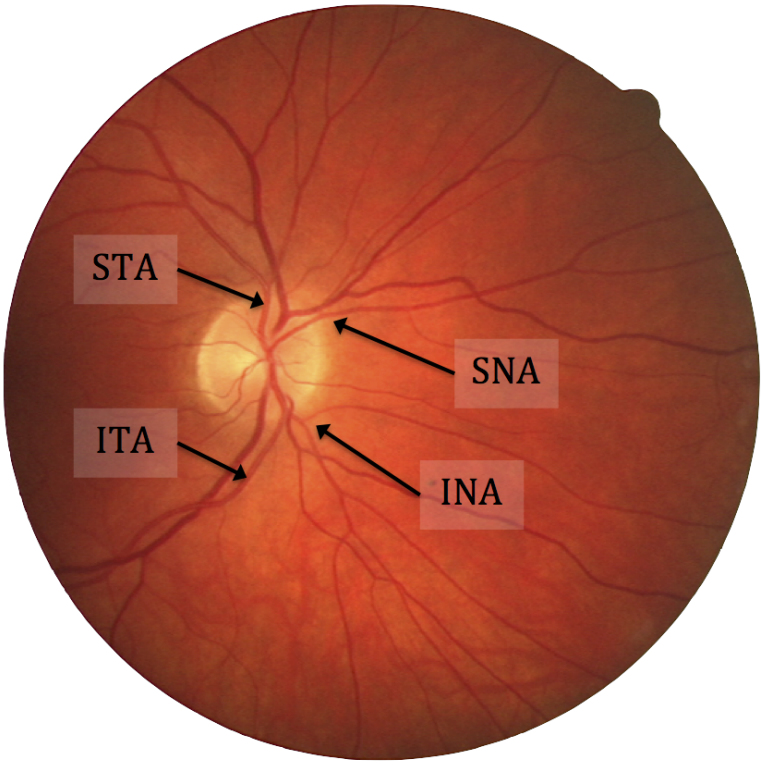


Figure 2.2: Arteries are distinguishable from veins due to being brighter red and narrower. Temporal and nasal refers to the relative position of the structures in the horizontal axis (temporal is closer to temporal side of the head, nasal to the nasal side). Temporal arteries and veins present an arch like structure, while nasal arteries and veins structure is closer to radial. The image is centered around the optic disk (not the macula). Image from the DRIVE dataset [1].

fact, a blind spot occurs, which means that any object focused on this region will not be perceived. The diameter of the optic disk is commonly used as a unit of length. For instance, it can be said that a given structure of the eye is three optic disks far from another.

Relatively to the vasculature, the retinal circulation (and nourishment) is possible by the perfusion of the central retinal artery (CRA) through the optic disk. The exception for this is the fovea, which is nourished through the choroid (due to being avascular). Four branches are originated from the CRA (see Figure 2.2):

1. Superior Temporal Artery (STA)
2. Inferior Temporal Artery (ITA)
3. Superior Nasal Artery (SNA)
4. Inferior Nasal Artery (INA)

Generally, standard retinal images are macula centered or, less commonly, centered around the optic disk.

2.2 Diabetic Retinopathy

Diabetic retinopathy (DR) is the most common diabetic eye disease, and can eventually lead to blindness. Data for the year 2012, shows that 5% of all the blindness in the world is caused by DR, and this percentage is much higher if age related eye disorders are excluded. In fact, it is considered the leading cause of blindness for people between 20 and 74 years old [7, 8]. The clinical importance of this data is paramount, especially considering the highly effective treatment options, in case of early diagnosis.

2.2.1 Cause and Evolution

DR is related with the systemic nature of diabetes, particularly the problematic glycemic control and hypertension and it is manifested by changes at the retinal microvascularity. These changes lead to a higher degree of vessel permeability, which in turn causes fluid leaks (edema). These leakages, depending on the exact location where they occur can degrade visual acuity (the extent to which this occurs is correlated with the distance to the macula). Other phenomena also occurs, which may have significant effects on the vision of the patient.

The changes associated with DR are, in general terms, cumulative and chronological, and can be summarized as follows:

1. Microaneurysm formation
2. Excessive vessel permeability
3. Vessel occlusion
4. Neo-vascularization and formation of fibrous tissue
5. Contraction of fibrous tissue

Due to the highly clinical value of these changes (as diagnostic means) a simple description will be presented for each of them, with special focus on the microaneurysms, since their detection will be the primary target of this project.

2.2.1.1 Microaneurysm Formation

As already mentioned, microaneurysms formation is the first evident (and visible) sign of DR. Histologically, microaneurysms present a loss of capillary cells (the perycites) on its walls. This leads to acellular capillaries, thus enabling the pouching of these capillaries, which originate the microaneurysms. The mechanisms that causes the decellularization are not well understood, but they include release of a vasoproliferative factor and an increase in capillary pressure [7].

Microaneurysms can be detected through a fundography examination. They appear as bright red dots, and have a diameter between 20 and 100 micrometers. Their small diameter and relatively poor contrast to the background (in color fundus images) makes them hard to detect. Due to their small diameter, precaution should be taken in the image acquisition to ensure that the image resolution is high enough for the microaneurysm to be present in at least some pixels.

Even though microaneurysms will be treated throughout this project as an early sign of DR, it should be noted that they may be associated with other pathologies or even not have any clinical significance, in cases where only a low number is present. However, a large number of microaneurysms or an increase in their numbers have been extensively linked to progression of DR [7]. This fact is part of the motivation for routine ophthalmoscopic screenings of diabetic patients.

2.2.1.2 Excessive Vessel Permeability

Microaneurysms formation often lead to excessive permeability, due to the changes that occur at the walls. This permeability often leads to macular edema - increase in thickness of the macula region due to fluid accumulation. Associated with the leaked fluids, there are often found lipid deposits, referred to as hard exudates. These deposits are visible in ophthalmoscopic images as yellow regions and often indicate permanent vision loss or degradation.

2.2.1.3 Vessel Occlusion

Vessel occlusion is one of the most problematic aspects of DR, and is generally associated with the changes that occur at the capillary walls. Occlusion may lead to haemorrhages and a deficient supply of blood to the macula (ischemia), thus compromising central vision.

2.2.1.4 Neo-vascularization

In a later stage, neo-vascularization occurs as a result of occluded blood vessels. New vessels are created due to the limited nourishment, but these vessels leak fluids easily (they have a very large permeability), thus aggravating the effects of DR. Fibrous tissue formation in the retina is also associated with neo-vascularization.

2.2.1.5 Contraction of Fibrous Tissue

As a consequence of the formation of fibrous tissue, retinal detachment may occur through the contraction of the newly created structures. This indicates a very serious situation,

requiring immediate care. It is visible as a continuous line separating portions of the fundus in fundography images.

2.2.2 Stages

Diabetic retinopathy is generally classified in two phases: non proliferative diabetic retinopathy (*NPDR*) and proliferative diabetic retinopathy (*PDR*). The two phases differ in the degree of changes verified in the retina microvascularity and, consequently, in the effects of DR in vision.

In the non proliferative phase, microaneurysms are generally the only evidence of retinopathy. Detection of DR in this phase poses a generally optimistic outcome. Proliferative DR is associated with neo-vascularization. This is an indication of severe DR, and the effects are no longer limited to the retina. In fact, neo-vascularization (and fibrous tissue formation) may also occur at the vitreous. The outcome for Proliferative DR is less optimistic, and permanent visual damage is often verified.

In this project the focus will be on *NPDR*, through the detection of microaneurysms.

2.2.3 Prevention and Treatment

Due to the progressive nature of DR, prevention and treatment can be analysed as interconnected aspects. The existent methods for addressing DR are relatively broad, and so will only be briefly summarized. However, a specific method, photocoagulation, has proved to be very robust, and will be explained in greater detail.

Due to the fact that DR is caused by poorly controlled physiological aspects, one approach to prevent it is exactly by maintaining these physiological parameters under control. One of the most important aspects is the glycemic control. As such, intensive insulin treatment is often prescribed to stabilize DR progression in at risk patients. Although less correlated, blood pressure has also shown to be an important factor in DR progression and strict blood pressure control is often recommended.

Another approach followed to treat DR is through the administration of pharmaceutical drugs. Generally, these drugs have antiangiogenic activity, thus inhibiting neovascularization. However, the efficiency of these drugs is limited and adoption is not widespread.

2.2.3.1 Photocoagulation

Photocoagulation is a technique where a laser is employed to cauterize (burn) blood vessels in the eye. Two types of photocoagulation are employed: focal photocoagulation and scatter photocoagulation. The former is applied directly to the microaneurysms or the

occluded vessel, while scatter photocoagulation cauterizes several regions in the retina, in defined intervals.

Although initially used to treat severe cases of DR, it is now more generally employed. In fact, several studies have shown [7] the high effectiveness of photocoagulation in preventing and even containing the proliferation of diabetic retinopathy, and, as such, it is extensively used whenever detection is confirmed.

Side effects are limited, but a small loss of visual field and visual acuity are often reported.

2.2.4 Diagnosis

Diagnosis of diabetic retinopathy is usually performed through an ophthalmoscopic examination, which allows the direct visualization of the eye fundus. This technique will be explained in greater detail in section 2.3.

Fluorescein angiography is also widely used for diagnosis purposes. This technique relies on fluorescein dye to show anomalous structures or lesions (such as microaneurysms or haemorrhages) in the retina. Sodium fluorescein is injected in the circulation and the emitted fluorescence allows the acquisition of the image of the retina. Despite presenting significant advantages over ophthalmoscopy in the detection of DR (for example, the easier detection of microaneurysms, as their contrast to the background is much higher than in ophthalmoscopic images), it also presents significant health risks, resulting in a mortality rate of 1:222 000 [9].

Other techniques used consist on standard assessments of vision parameters such as acuity tests, which allow to easily and noninvasively detect deficiencies in the visual function.

2.2.5 Screening Practices

Due to the asymptomatic nature of the early stages of DR, it is difficult to perform a timely diagnosis of the disease. Due to this reason, adequate and periodic ophthalmoscopic screening is recommended.

Data from the Wisconsin Epidemiological Study of Diabetic Retinopathy (WESDR) suggest an yearly screening procedure for type 2 diabetes patients. Other data suggest larger intervals (three years) for patients with no retinopathy signs, but also emphasize yearly diagnosis for those which evidentiate any sign.

These guidelines are often disregarded in current clinical practice due to the difficulties inherent to performing large scale screening programs. In fact, most patients are screened either due to being diabetic for a long period (more than ten years), presenting systematic risks or showing repeated risk behaviours (such as performing deficient insulin administration).

These deficient screening practises and their consequences are the main motivating factor for this project.

2.3 Ophthalmoscopy

Ophthalmoscopy is a technique that allows the visualization of the fundus of the eye. As previously explained, ophthalmoscopy (also referred as fundoscopy) is the most used method in the diagnosis of early diabetic retinopathy.

The principle of the ophthalmoscopic exam, as stated in [10] is that *the eye can be made luminous if the axis of illumination directed towards a person's eye and the line of vision of the observer are coincident*. This means that in order to visualize the fundus of the eye, besides the observer, the illumination also needs to be aligned with the structure to be visualized, similarly to the representation in Figure 2.3.

The ophthalmoscopy of the eye is commonly divided in two different approaches: direct ophthalmoscopy and indirect ophthalmoscopy.

2.3.1 Direct Ophthalmoscopy

The first direct ophthalmoscope was invented in 1846 by Hermann von Helmholtz [11]. This innovative device solved a constraint that arose from the aforementioned ophthalmoscopic principle: the fundus can only be visualized where the illumination and observations areas overlap.

Discounting possible optical abnormalities, this situation only happens when the light source and the observer's pupil are aligned. The pupil generally appears black exactly because a light source rarely is aligned with an observer. A good example of an exception is in camera photographies with flash, where the very strong flash light is aligned (approximately) with the objective, leading to the well known *red eyes* effect, where the pupil has a red color instead of the normal black - the red color is due to the reflectance of light in the eye fundus (which has a reddish color).

Different techniques can be employed to achieve the optical alignment between the observer and the light source [11]. Hand held direct ophthalmoscopes typically use a design such as represented in Figure 2.4, where the light source is directed to a mirror, striking on the lower half portion of the pupil. The upper portion is used to observe the reflection, and thus it is possible to observe the eye fundus. Other designs can also be used, depending on the application.

2.3.2 Indirect Ophthalmoscopy

Due to the intrinsic physics of the direct ophthalmoscopy method, only a small area of the fundus (small Field of View - FOV) can be observed. The reason for this is that the light

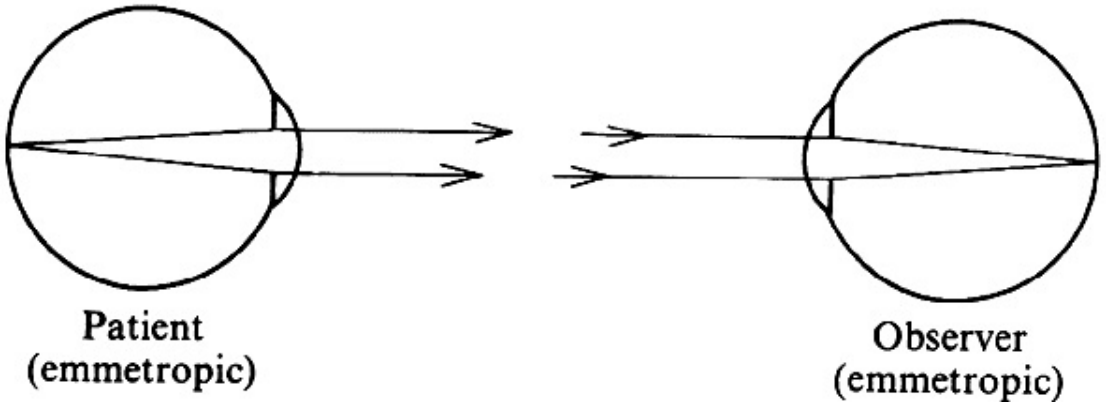


Figure 2.3: Representation of the ophthalmoscopic principle, considering that both the patient and the observer have emmetropic eyes (the images are focused precisely on the retina). Image from [11]

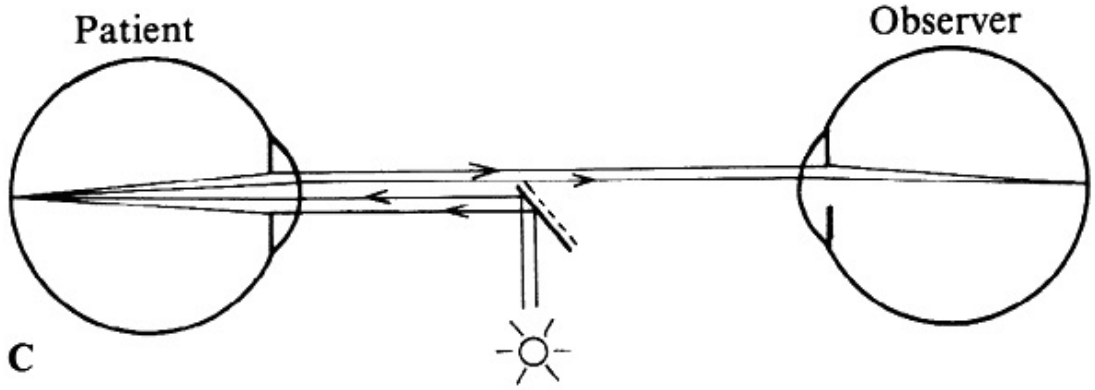


Figure 2.4: Illumination method used in current direct ophthalmoscopes, with a prism. Image from [11].

rays need to pass the patient pupil, the reflection must also pass the patient pupil and they finally have to pass the examiner's pupil. These constraints strongly affect the available FOV.

In 1852, Ruete proposed a different approach for the ophthalmoscopic examination, using an intermediate lens [11]. The objective of this lens is to condense rays reflected on the fundus of the patient and that passed it's pupil, but which wouldn't (on direct ophthalmoscopy) strike on the observer's pupil. The principle of indirect ophthalmoscopy is represented in Figure 2.5.

The distance of the patient's eye to the lens is similar to the focal distance of that lens. On the other hand, the FOV is determined by the product of the lens diameter by its dioptric power ($FOV = LensDiameter * Dioptricpower$).

One disadvantage commonly mentioned against indirect ophthalmoscopy is the relative difficulty of the technique [11]. This is due to the large distance of the patient pupil to the lens, making the alignment with the light source and the examiner's own pupil

complicated.

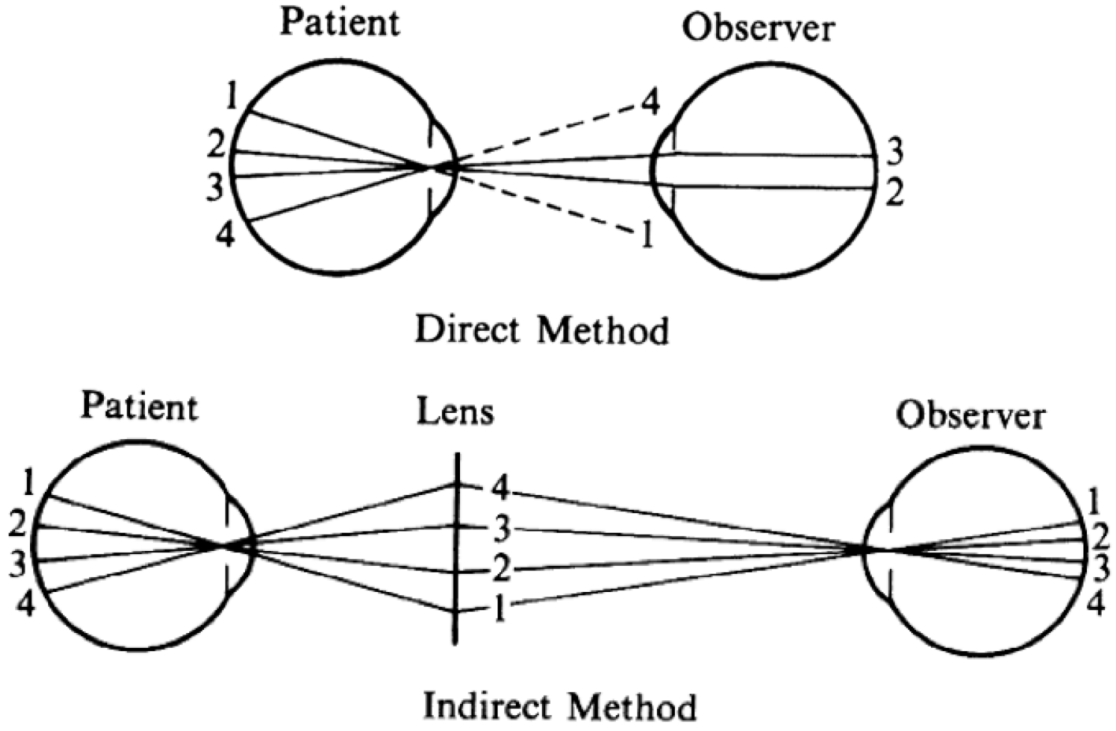


Figure 2.5: Comparison between direct ophthalmoscopy and indirect ophthalmoscopy. Image from [11]

2.3.3 Fundus Photography

Fundus photography can be seen as a special type of ophthalmoscopy, allowing not only the observation of the fundus of the eye, but also supporting proper documenting and follow up. Systems for fundus photography employ the indirect ophthalmoscopy principles, replacing the observer by a camera.

Some fundus cameras support infrared (IR) as a light source. Use of IR has some advantages over visible light, as it spares the patient from the uncomfortable strong visible light used in traditional ophthalmoscopy. IR light has some other interesting properties, such as the ability to better penetrate obfuscated media (important for imaging cataract patients, for instance).

Due to the enclosed construction of fundus cameras, significant improvements (in terms of illumination, FOV and general image quality) over traditional techniques are possible.

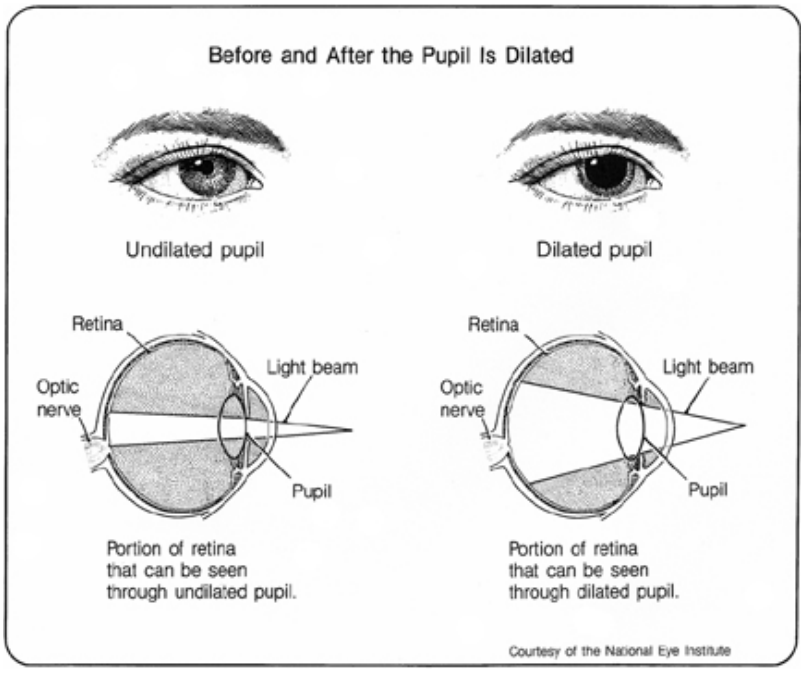


Figure 2.6: Effect of pupil dilation. Obtained from the National Eye Institute.

2.3.4 Pupil Dilation

From the fundamental principle of ophthalmoscopy, it can be observed that one of the limitations of the eye examination will be the size of the patient pupil, which will serve as an aperture.

In order to obtain a larger FOV of the fundus of the eye, pupil dilation - also referred to as *mydriasis* - will be useful (please refer to Figure 2.6). This can be achieved by the use of physiological solutions such as tropicamide or phenylephrine. These solutions stimulate the iris dilator muscle, leading to the dilation of the pupil.

Besides an increase in the obtained FOV, the retina background tends to be more uniform in retinal images acquired with pupil dilation, leading to higher quality images.

2.3.5 Supporting Software

Some software is available for supporting the task of performing an ophthalmoscopic examination and managing the acquired images.

A particularly important supporting task is the stitching of several contiguous images. This allows an increase in the retina area available for evaluation. *Ginkgo CADx Retina* and *Dual Align i2k Retina* are examples of commercial software tools that perform this function.

Image enhancement is also frequently used to facilitate the detection of abnormal structures. Transformations in the red channel of the image are frequently used, in order to improve contrast. Again, *Ginkgo CADx Retina* also supports this procedure.

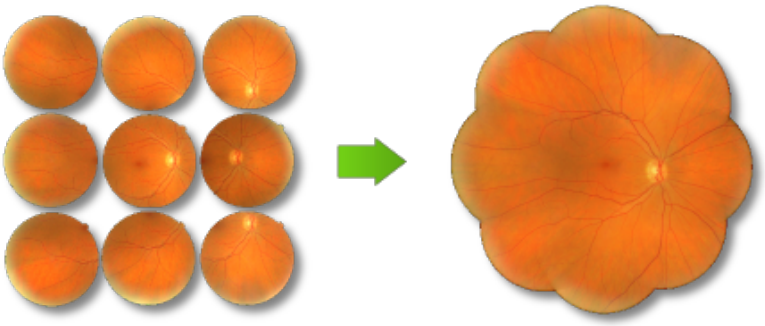


Figure 2.7: Example of retinal image stitching, performed by *Ginkgo CADx Retina*.

Desktop ophthalmoscopic devices usually use their own software for control of parameters such as the light intensity, exposure, and zoom. Basic image analysis capabilities are also generally covered. Information systems tend to be included, allowing the proper registration of patient data with the acquisition. *Canon Retinal Imaging Control Software* is an example of this software, for *Canon* own retinal cameras.

2.3.5.1 Automated Diagnosis Software

Besides the performance of supporting tasks such as image and patient data management, and (relatively) simple image processing, some commercial software to perform automated diagnosis exist.

iGrading, commercialized by *Shil* and *IDx-DR* are examples of this type of software. They provide interfaces to interact with different patient management tools and using an automatic algorithm, detect early signs of DR.

RetmarkerDR, commercialized by *Retmarker* is a more comprehensive solution, comprising acquisition (a fundus camera is included), software for interfacing with the fundus camera, and also remote image diagnosis (where the automated diagnosis is performed). DR progression is tracked, and comparative analysis of images is used to better predict it. In cases where the diagnosis algorithm detects significant risk, the case is sent to a triage technician, for a second analysis.

All mentioned software is certified as a Medical Device under European Union law (*Class I* for *iGrading*, *Class IIa* for *IDx-DR* and *RetmarkerDR*), necessary for commercialization of automated diagnosis software in the European Union.

2.4 Related Work

Due to the epidemiological characteristics related to DR and the importance of early diagnosis, several different methods were proposed in the literature to tackle the problem of

DR (and, more specifically, microaneurysms) detection. A succinct overview of the most popular will be presented in this section.

Despite the relatively high variance of the existent methods, they generally follow a similar high level approach.In fact, most methods divide the problem of detection into two phases: candidate extraction and classification [12].

Candidate extraction includes several image processing steps, including adequate image pre-processing to reduce noise and improve contrast, and some form of lesion enhancement followed by thresholding, which should output the candidate regions.

The second stage consists in the classification of the candidate regions as true (in case the candidate is a microaneurysm) or false. This requires a previous training step, where the classification algorithm will *learn* how to detect microaneurysms by building a model based on correctly labelled microaneurysm regions. In the training step, publicly available datasets with credible ground truth information are often used to extract the relevant features for classification. In this project, a public dataset with ophthalmoscopic images was also used to assess the performance of the developed algorithm, as explained in section 3.3.

Several methods [13, 14] use angiography images rather than ophthalmoscopy to perform automatic detection of microaneurysms. The rationale beyond this is that microaneurysms are relatively difficult to detect in ophthalmoscopy, as they are often indistinguishable from other existent structures or lesions. In angiography, however, microaneurysms are very hyperfluorescent [7] (please refer to Figure 2.8), and thus are easier to detect. Nonetheless, the health hazards associated with this technique (as mentioned in 2.2.4) compromise its broad use for screening purposes.

In ophthalmoscopic images of the fundus, generally the operations are performed in the green channel of the original RGB image. This is a result of the higher contrast between the microaneurysms and the background structures in the green channel - the red channel is generally saturated and the blue channel typically is not very informative. This is illustrated in annex A.1.

Another almost ubiquitous operation is the removal of shade artefacts (for example, in the region of the macula) by means of minimization of the background intensity variation. Blood vessel segmentation is also commonly performed in order to exclude false positives - while microaneurysms occurs at the capillaries, these are not visible in fundus images. Several methods are already developed for the task of vessel segmentation and some of those already achieve performance similar to human graders [15].

Three different general approaches will be considered, similarly to the work in [16]:

- Morphological Processing
- Neural Networks
- Classification

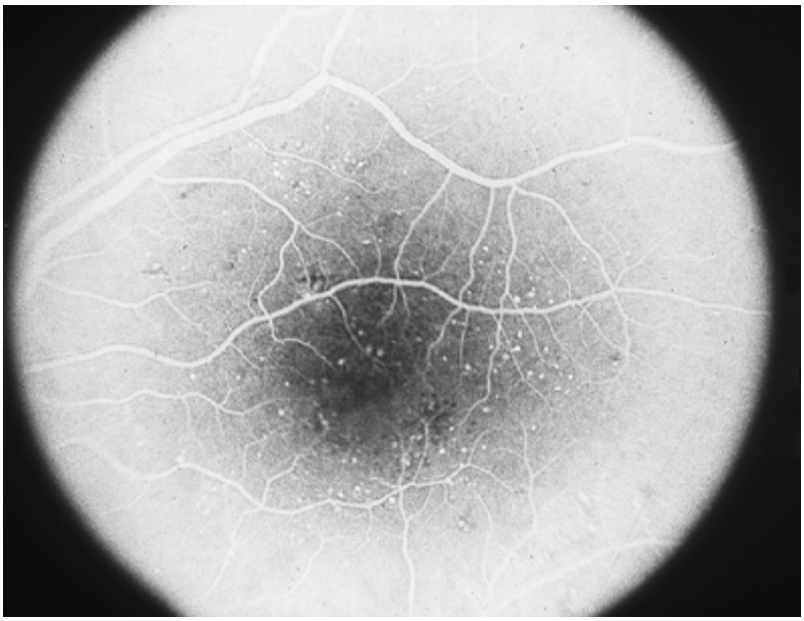


Figure 2.8: Angiography image of the retina. Microaneurysms correspond to the bright spots in the macular region. Image from [7].

It should be noted that these divisions are not totally precise, and in fact several methods fit more than one of these categories, as they may include, for example, a morphological and a classification step. Still, the innovative component in the most influential methods generally belongs to one of these categories, and this is the rationale for this sectioning.

The presented results are on a per image basis - a true positive is considered if MAs are detected on an image which in fact presents them. In contrast, the evaluation of the developed algorithm (please refer to section 6.1.3) will be on a per lesion basis - a true positive is only considered when a microaneurysm and its location are correctly detected.

2.4.1 Morphological Processing

Morphological processing is the most common method for DR detection. These methods are associated with the performance of morphological operations on the retinal images in order to enhance possible microaneurysm regions, which will then be candidates for a posterior classification step. Some methods avoid classification and use simpler processes instead to obtain the final decision.

A particularly influential method has been proposed by *Spencer 1996*, which employed a top hat transformation to segment candidate regions. The purpose of this transformation was to discriminate elongated structures (the vessels) in favour of circular isolated red lesions, which fit the microaneurysm candidate profile. This was accomplished

by using twelve rotated structuring elements (with size greater than the expected microaneurysm): taking a map of the maximum pixel values of all of the top hat results and subsequently thresholding that map, the vasculature segmentation is obtained. It is then subtracted to the already pre-processed image, thus removing the vessels while maintaining possible microaneurysm regions.

The method suggested in *Niemeijer 2005* [17] further improves this solution by introducing a pixel based classification for the detection system and a large number of features for performing the classification. The first improvement addresses a substantial shortcoming of the method in *Spencer 1996* [18], as the size of the structuring element (SE) requires a compromise between a good vessel segmentation (small SE) and a low false negative rate for microaneurysms (large SE, thus avoiding the disregard of small structures). This first step classifies individual pixels and combines the results with the method by *Spencer 1996* in order to obtain candidates for a posterior k-nearest neighbour classifier. A larger number of features for the classification were also used, consisting in shape features, pixel intensity features, Gaussian filter-bank outputs and color features.

The method has shown very good performance, registering a sensitivity of 100% and a specificity of 87%, for a dataset comprising 100 images.

2.4.2 Neural Networks

Due to their universal function approximator property, neural networks have also been popular for tackling the problem of microaneurysm detection in fundoscopy images.

Garcia 2008 [19] described a method to detect microaneurysms which employ a multilayer perceptron. A set of 29 features related to the color and shape were used for the training and testing. A sensitivity of 86.1% was achieved using 50 images for testing.

This approach has been further improved in *Garcia 2010* [20] where four neural networks based classifiers were used: a MLP (Multilayer Perceptron), RBF (Radial Basis Function), SVM (Support Vector Machine) and an hybrid based on the voting of all the previous. Similar features were obtained.

2.4.3 Classification

A wide range of solutions are based on a classification approach. A quick analysis of some of the most important methods will be presented.

A method based on the previously described, by *Niemeijer 2005* [17], has been proposed by *Balasubramanian 2008* [21], which deals with some of the performance issues found. This method employs a different approach to the candidate extraction step, using an Automatic Seed Generation procedure instead of the top hat transform. Classification is performed by a hybrid of two classifiers, k Nearest Neighbour and Gaussian Mixture

Models. A sensitivity of 87% and a specificity of 95.53% were achieved using a dataset of 63 images.

Hatanaka 2012 [22] proposed a method for detecting microaneurysms using a rule based classifier as well as an artificial neural network. A large number of features (126) were obtained from the image after being processed by a double ring filter and 28 components were selected using Principal Component Analysis. A true positive rate of 68% was obtained using the publicly available ROC dataset.

Chapter 3

Eye Fundus Acquisition and Image Analysis

3.1 Ophthalmoscopic Adapter

As the main objective of this project is the development of an integrated solution (comprising acquisition and detection) for the diagnosis of diabetic retinopathy using a smartphone, an ophthalmoscopic adapter that allows image acquisition of the eye fundus is required. Due to this fact, an analysis of the existent ophthalmoscopic adapters has been performed. The solutions found should conform to a set of technical requisites:

- Field of View (FOV) greater then 20^{o}
- Sufficient magnification for the resolution of microaneurysms
- Variable focus (preferentially)
- Independent illumination (preferentially)

Field of View is a very important specification, as it will define the retina area extent that will be available for visualization. Traditional direct ophthalmoscopes are particularly limited in this aspect, mostly presenting FOVs around 5^{o}. A low FOV hinders the entire screening process, as even if multiple images can be taken (to examine a larger area of the fundus), this process is not practical. An ideal solution should present a FOV around 40^{o}, which is a specification typical for fundus cameras, from which most literature data sets are obtained. However, this is not realistically attainable for mobile solutions. Furthermore, non mydriatic (without pupil dilation) solutions are preferred, due to some side effects of the dilation, and this would further hinder the available FOV. A 20^{o} FOV is an acceptable value, as it already allows the inclusion of significant number of structures around the optic disk or macula.

The chosen adapter should have sufficient magnification to allow the resolution of microaneurysms. This specification is harder to enforce due to the fact that most solutions do not directly specify their magnification. However, as this condition is associated with the optical properties of the device, it can be related with the specified FOV.

Characteristics such as variable focus and independent illumination are preferred but not required. These specifications could help obtain higher quality images, as well as ensuring a more homogeneous acquisition with different smartphone cameras.

Search for suitable adapters was performed by consulting the academic literature as well as the intellectual property protecting those inventions. Four different solutions were found to be viable:

- PEEK (Portable Eye Examination Kit)
- Welch Allyn PanOptic Ophthalmoscope + iExaminer
- Ocular CellScope
- Indirect Ophthalmoscopy

Table 3.1 presents an overview over the specifications of the different solutions.

	PEEK (Portable Eye Examination Kit)	**Welch Allyn PanOptic Ophtalmoscope+ iExaminer**	**Ocular CellScope**	**Indirect Ophthalmoscopy**
Field of View	25º	25º	55º	40º
Illumination	LED/Flash smartphone	Adjustable illumination is included	Illumination included	LED/Flash smartphone
Device Compatibility	Compatible with most standard smartphones	iPhone 4/4s	Any Smartphone	Any Smartphone
Validation	CE Certification	CE and FDA certification	Currently being tested in India and Thailand	Uses the established indirect ophthalmoscopy technique
Pupil Dilation	Recommended (increases FOV)	Recommended (increases FOV)	Recommended (increases FOV)	Required
Price	Not yet available (estimated to 75€)	826€(ophthalmoscope) + 67€(adapter)	Not yet available (price of components $900)	VOLK Bio Lens 20D: $300 (standard Indirect Ophthalmoscopy Lens)

Table 3.1: Comparison of the different adapters specifications.

3.1.1 PEEK

PEEK (Portable Eye Examination Kit) has been developed by *PeekVision*, an UK-based startup. The device has been initially tested with a *Samsung Galaxy S3* smartphone. The adapter was made using 3D printing for the casing and a laser cutter for the optics [23].

PEEK doesn't integrate autonomous illumination, instead relying on the smartphone own LED flash as the ophthalmoscopic illumination source. Even though this simplifies the device design, it also increases acquisition heterogeneity across different smartphones.

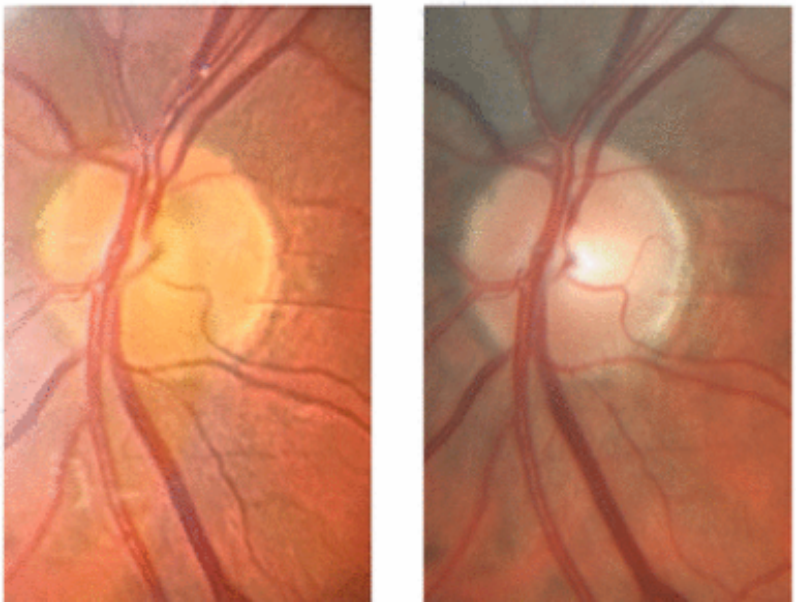


Figure 3.1: Image acquired by PEEK (left) and image acquired by a conventional fundoscopy device (right, *Topcon TRC-50EX*).

It also doesn't possess focus capabilities, thus limiting the focusing available to the one performed by the smartphone camera.

The FOV value of the PEEK adapter is around 25º, much larger than the FOV for traditional methods of direct ophthalmoscopy. Validation of this device against traditional fundoscopy cameras is currently under way in Kenya (please refer to Figure 3.1). Medical certification was already obtained for the European Union (but not for the USA market).

Even though this solution presents several advantages, it is not yet commercially available. *PeekVision* set October 2015 as the date for the beginning of commercialization, which obviously does not comply with the timeline for this project. Contact with *PeekVision* was made to try to obtain a test version of PEEK, but the limited number of available devices did not allow this possibility.

3.1.2 Welch Allyn PanOptic Ophthalmoscope + iExaminer

Welch Allyn is one of the main manufacturers of clinical ophthalmoscopes and PanOptic is their high end model for portable ophthalmoscopes. *Welch Allyn* also developed an adapter for this device that allows it to be used with an *iPhone* smartphone - but only with the model 4 and 4s (see Figure 3.2). However, as this adapter does not have any optic function (it only supports the ophthalmoscope in the smartphone) developing a similar adapter for an Android smartphone is relatively easy.

The greatest advantage of this option is the superior optics and the integrated illumination system, which allows greater magnification and better overall image quality. The PanOptic allows independent focus control (in a range of -20.0D to +20.0D).

Figure 3.2: Welch Allyn PanOptic Ophthalmoscope with an iPhone smartphone.

An iPhone application is available to perform the acquisition and management of the fundus images. It captures a video sequence from the camera for a short period of time and then allows the user to choose, from the several frames, one that is representative and saves that image with the adequate patient data.

3.1.3 Ocular CellScope

Ocular CellScope is a device developed around 2007 in the University of Berkeley, from a general mobile microscope solution. This device is currently still in development, but the specifications are interesting, especially the FOV (around 55^{o}). However, there is no defined timeline for the release of a commercial version of this product.

3.1.4 Indirect Ophthalmoscopy

Significant research has been performed for the application of the technique of indirect ophthalmoscopy (see section 2.3.2) to smartphones. This technique uses a lens, generally at a fixed distance from the smartphone camera, to perform the acquisition of the eye fundus.

While good results are attainable (and a large FOV can be captured), this approach has two significant disadvantages: the first is the fact that the technique is rather difficult to efficiently perform (the acquisition would not be possible in a self assessment context, or even for a general practitioner); the second is that this method requires pupil dilation, which has some side effects and cannot be performed without a proper specialist.

3.1.5 Adapter Choice

Even though PEEK Adapter has several interesting specifications and a lower price than the other adapters (making it more appropriate for self examination), it is unavailable. As such, after considering the advantages and disadvantages of the available adapters, the *Welch Allyn PanOptic Ophthalmoscope + iExaminer* was chosen to be used in this project.

3.2 Acquisition Protocol

In order to standardize the acquisition conditions for development purposes, an acquisition protocol was developed. This protocol, described in appendix B, defines the interaction with the examined person, the adjustment of the ophthalmoscope settings and the proper technique to use during the acquisition.

3.3 Datasets

One of the objectives of this project is to develop a mobile solution capable of detecting DR signs (specifically microaneurysms) through the analysis of retinal images taken using a smartphone. Even if this is the ultimate approach, in order to develop the method, validate it and compare it to other state of the art approaches, it should be tested on public datasets with available ground truth.

Two different datasets were used in the development of this project: the DRIVE dataset (Digital Retinal Images for Vessel Extraction) and the ROC dataset (Retinopathy Online Challenge).

3.3.1 DRIVE

The DRIVE (Digital Retinal Images for Vessel Extraction) dataset was obtained from a screening program for diabetic retinopathy, performed in the Netherlands [1]. From a screening population of 400 diabetic subjects, 40 photographs were randomly selected and are present in the dataset.

The images were acquired using a Canon CR5 camera with a 45^{o} field of view. Images have a resolution of 768 by 584 pixels and 8 bits per color plane.

The 40 images were divided into a training and test set, both sets containing 20 images. For the training images, a manual segmentation of the vasculature is the ground truth. For the test cases, two manual segmentations are available, one used as gold standard, and the other one can be used to compare computer generated segmentations with those of an independent observer.

This dataset can be used to test/train algorithms for vessel segmentation, being useful for the removal of false positives on the MA detection method and also for performing image stitching (merging images from different regions of the fundus in order to obtain a larger area).

3.3.2 ROC

The ROC (Retinopathy Online Challenge) dataset presents a distinctive philosophy to the DRIVE dataset, as the ground truth is only available for a subset of the images. This is one of its main purposes, as it allows the evaluation of the output of a diabetic retinopathy CAD algorithm in a uniform manner, and comparison between different algorithms. Furthermore, this dataset is specific for the detection of microaneurysms and haemorrhages, and their location in each image is the ground truth [24].

The dataset consists in 100 images of digital color fundus photographs from an original larger dataset of 150 000. These images were obtained in different conditions and so are heterogeneous (44 images have a resolution of 768 by 576 pixels, 9 have a resolution of 1058 by 1061 pixels and 47 have a resolution of 1389 by 1383 pixels). 50 of the images and their respective ground truth compose the training data and the other 50 compose the test data.

The methodology for the elaboration of the ground truth differs for the training and test datasets. Four experts annotated both datasets, but for the training dataset a microaneurysm was considered even if only one of those human raters marked it. For the test dataset, the annotation of one randomly selected expert was discarded (was used to assess human performance) and a microaneurysm was marked if there was a consensus for its location between at least two of the remaining three experts.

3.4 OpenCV

OpenCV (Open Source Computer Vision) is a library of C and C++ functions aimed at computer vision applications, more specifically for real time computer vision. A wide range of image analysis and processing methods (for filtering, segmentation, geometric transformations and feature detection) are available and well optimized.

Common applications of this library include face recognition, motion detection, augmented reality, robotic navigation support and medical diagnosis.

3.4.1 OpenCV with Android

Even though the OpenCV functions are implemented in C++, Java wrappers are available, and so it is possible to easily port developed code. Furthermore, native C++ code may be

used, through the *Java Native Interface* (JNI). Due to a relatively large overhead in calling each wrapped OpenCV function, C++ code should be preferred in repetitive calls.

As a result of the relatively large size of the OpenCV libraries, a third party application, *OpenCV Manager*, needs to be installed. This can be avoided, but the included libraries would increase the developed application size by ~ 20 MB.

3.5 Image Stitching

Due to the relatively constrained FOV available through the use of the PanOptic ophthalmoscope, the number of retinal structures that can be analysed in a single image is limited. As mentioned in 3.1, while fundus cameras generally present a FOV greater than 40^{o}, the PanOptic ophthalmoscope maximum FOV is around 25^{o}. This problem can be addressed by taking multiple overlapping images of the retina and merging these results to form a larger FOV.

The problem of merging multiple images to form a larger one is one of the main commercial applications of image processing (for example, to build panoramic images, as in figure 3.3) and a large body of work exists on this subject. On a general level, stitching two (or more) images requires two steps [25]:

Registration Finding the optimal alignment between the images to be stitched.

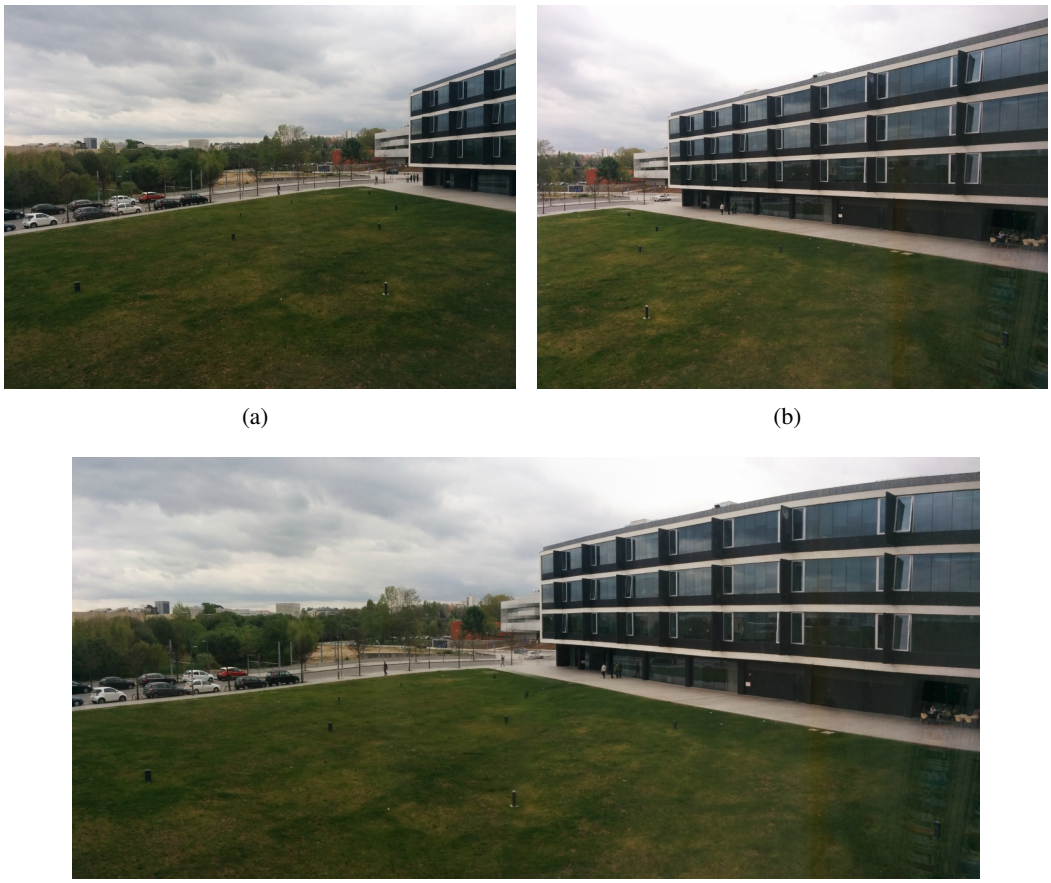
Compositing Given the alignment, merge the images, performing bundle adjustment, finding the optimal seam mask to merge the images and homogenizing the final result.

3.5.1 Registration

The alignment (or registration) between different overlapping images can be expressed by means of geometric transformations. Possible transformations include translation, rotation, scaling and the more general affine and perspective transformations. Transformations are represented by matrices: mapping an image into another corresponds to multiplying the original coordinate system of one of the input images by the appropriate transformation matrix (generally denoted by the letter H) [25].

3.5.1.1 Geometrical transformation models

As explained, different models can be used to parameterize the transformation between images. These different models offer different degrees of freedom and can describe with more or less flexibility different transformations.



(a) (b)

(c)

Figure 3.3: Example of image stitching. (a,b) input images; (c) result of stitching.

3.5.1.2 Translation

The transformation matrix for a translation (t_x, t_y) is represented by:

$$H = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix}$$

Translation transformations offer two degrees of freedom (along the x and y axis), and they preserve the orientation, lengths, angles, parallelism and straight lines in the mapped image.

3.5.1.3 Euclidean (Rotation and Translation)

Euclidean transformations account, besides translation, for rotation around a given angle θ. The transformation matrix is represented by:

$$H = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \end{bmatrix}$$

Euclidean transformations offer three degrees of freedom and preserve lengths, angles, parallelism and straight lines in the mapped image.

3.5.1.4 Similarity

Similarity transformations also allow scaling of the images. The transformation matrix is represented by:

$$H = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix}$$

Similarity transformations present four degrees of freedom and they preserve angles, parallelism and straight lines.

3.5.1.5 Affine

Affine transformations are more general and may express six degrees of freedom. They only preserve parallelism and straight lines and their matrix representation is:

$$H = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix}$$

3.5.1.6 Perspective

Perspective transformations are the most general transformations and present eight degrees of freedom, only preserving straight lines. The transformation matrix has size 3×3 and the mapping performed by this transformation to the coordinates x and y is:

$$x' = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + h_{22}}$$

$$y' = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}}$$

3.5.2 Alignment Methods

In order to find the transformation matrix that better describes the alignment between two images, two different methods are commonly used: feature based alignment and direct alignment.

Both methods require a transformation model to be initially specified and will then estimate the model parameters. Choosing a suitable transformation model is important: while general transformation models (like an affine or perspective transform) may be more flexible, as they account for higher variations, their computation is often more computationally expensive and/or less accurate than less flexible models, if the real transformation is simple.

3.5.2.1 Feature based alignment

One of the approaches to perform registration of two overlapping images is using corresponding points, which are points that represent the same region/object in both images. These points are automatically extracted and described by appropriate methods. Using the points known correspondence it is possible to calculate the geometric transformation between images [25].

Specific methods are used for each step of the alignment process and they will be briefly reviewed in the next sections.

Feature Detectors

Feature detectors are the algorithms that automatically detect potentially interest points (also known as features) in each image. Interest points are points (or small regions) that are considered descriptive of the image. For example, in figure 3.4(a), interest points can be the ones relative to the letters on the building side or along the edge of that building.

Generally those interest points are located in corners, where an intensity change is verified across multiple directions. In fact, several feature detectors [26, 27] use a variation of corner detectors - filters that have high response when an intensity change occurs in more than one direction - to obtain the location of the interest points. The result of a feature extractor on the mentioned image is shown in figure 3.4(b).

Feature Descriptor Extractors

Feature descriptor extractors are used to provide relevant descriptors for each identified interest point. These algorithms use information about the point and its neighbourhood to compute the descriptor vector. A significant challenge is to provide scale and orientation invariant descriptors, as scale and orientation assumptions generally can not be made when trying to align different images.

(a) (b)

Figure 3.4: Interest point detection. (a) Original Image; (b) Interest points detected by ORB in the original image (represented by coloured circles).

Some popular examples of feature detectors and descriptor extractors algorithms are SIFT (Scale Invariant Feature Transform) [28], SURF (Speeded Up Robust Features) [26] and ORB (Oriented FAST and Robust BRIEF) [27]. The existent algorithms differ on their computational performance, on the number of extracted keypoints and on the number and type of features (descriptors) obtained. Furthermore, some of these methods are protected by intellectual property rights (specifically SIFT and SURF), which may deter their use in some applications.

Feature detectors and descriptor extractors can be used separately - for instance, it is possible to use the SIFT feature detector to extract the interest points and the SURF feature descriptor extractor to generate their descriptor vectors.

Matching Interest Points

After obtaining the interest points and their descriptors, they should be matched. A possible approach is to use a distance metric (euclidean distance, for example) between descriptor vectors and classify as a match two points if the distance between their feature vectors is inferior to a specified threshold.

Comparing every interest point with each other has time complexity $O(N^2)$ and since the number of extracted points is generally large, an optimized method should be used to perform this matching. An optimization of k-d trees can be used to find approximate nearest neighbors between descriptor vectors faster [29]. An example of matched interest points across two images is presented in figure 3.5.

Figure 3.5: Corresponding features in two overlapping images. Every detected interest point is represented by a coloured circle and correspondences between images are represented by lines.

RANSAC

After matching interest points to find correspondences between images, the transformation matrix must be estimated, according to a specified transformation model (as previously reviewed in section 3.5.1.1).

Different models require different number of points for their parameters to be estimated. Translation transformations require one matching point, euclidean transformations requires two points, similarity and affine transformations require three points and perspective transformations require four points.

However, the number of matching features is generally higher than the number of matching points required and mismatches may also occur. This problem is generally addressed by using RANSAC (RANdom SAmple Consensus). RANSAC iteratively chooses only a sufficient number of correspondent points (one to four, depending on the transformation model) and computes the respective alignment [30]. Then, it is estimated how many of the remaining points are inliers - i.e., how many will be correctly estimated using the computed alignment, according to their already known correspondence. After all the iterations, the parameters which generate more inliers are chosen to describe the transformation.

The probability of RANSAC finding the optimal alignment, after S iterations, for a p probability of a given correspondence being a true positive and for a transformation model that requires k matching points is [30]:

$$P = 1 - (1 - p^k)^S$$

For a problem where the probability of each match being a true match is only 50% and where an euclidean model is considered (two matching points needed, $k = 2$), the probability of the alignment being correct after 20 iterations is 99.68%. This illustrates that

even for a difficult case where the percentage of mismatches is very large, the RANSAC algorithm provides a robust estimation of the alignment. This probability equation also illustrates why choosing a more general model would complicate the computation of the alignment: if a perspective model was considered (four matching points required, $k = 4$), the probability of the alignment being correct after the same number of iterations would be 72.49%.

3.5.2.2 Direct Alignment

Direct alignment is a different alignment approach that uses all the image pixels to estimate the optimal registration. Different parameters (relative to a chosen transformation model) are tested and pixel to pixel agreement is calculated. The objective function is to maximize this agreement (or alternatively, minimize the disagreement).

The disagreement between pixels is measured by an error metric. Possible error metrics are Sum of Squared Differences (SSD) or Sum of Absolute Differences (SAD):

$$SSD = \sum_i [I_1(i) - I_{2T}(i)]^2$$

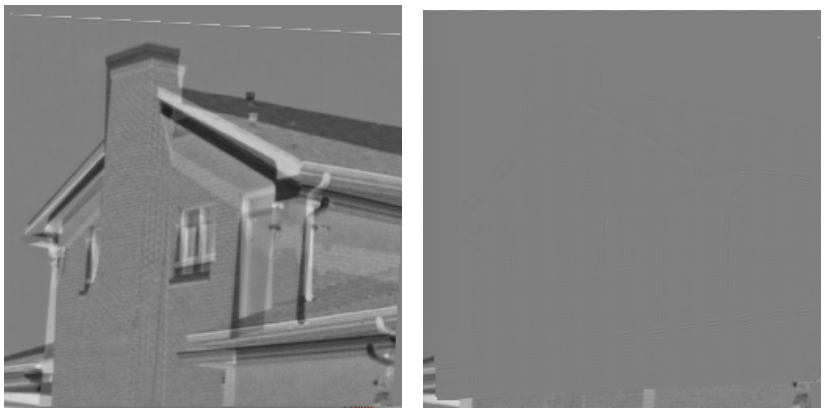
$$SAD = \sum_i |I_1(i) - I_{2T}(i)|$$

As specified, the objective of direct alignment is to minimize the error metric and, as such, a suitable method should be used. A brute force approach, consisting in testing all possible alignment would be infeasible to all but severely constrained transformations (for instance, if it is known that a translation would be smaller than 5 pixels).

The method typically used to perform direct alignment is hierarchical motion estimation [25]. In this method, a set of different resolutions of the original images is generated, and the optimal registration for the pair of images with the lowest resolution can then be computed efficiently. The result is used as an approximation for the next pair (of higher resolution), and the process is repeated for all the increasing resolutions. The calculated alignment can be further improved, being possible to obtain subpixel accuracy, through different methods [31, 32]. Figure 3.6 illustrates the initial difference between two images and the final difference, using the computed alignment.

3.5.2.3 Feature based alignment vs Direct Alignment

The two approaches to perform image alignment present different characteristics. While feature based alignment is generally considered less accurate, it is more robust in finding the alignment when only a small overlap exists. Direct alignment is considered more accurate (subpixel accuracy is possible, which is useful for advanced applications, such



(a) (b)

Figure 3.6: Example of image difference before and after direct alignment. (a) Before direct alignment; (b) After alignment. An affine transformation model was used to compute the alignment. From [33].

as multi image denoising), but it often fails to find an alignment when the input images have only a small overlap.

3.5.3 Compositing

After the alignment is computed, the input images need to be composed in order to obtain the larger stitched image. This generally involves a series of steps, which include image matching, bundle adjustment, seam estimation and image blending.

Image matching corresponds to finding how multiple images can be aligned with each other, given a set of input pairwise alignments. A possible approach is to construct a graph of connected images, where the nodes are the input images and the edges between them correspond to the existence of an alignment. The graph does not necessarily need to conform to the computed alignments, as some alignments can be discarded in case they are not sufficiently informative (for example, two images align but are similar) or in case their alignment does not conform with a majority of other alignments. The nodes (images) in the largest connected component of the graph are the ones picked to compose the final stitched image.

Bundle adjustment estimates camera parameters (rotation, focal length) by performing adjustments in order to account for small discrepancies between different alignments.

Seam estimation consists in finding the optimal regions of each image that will contribute to the final image. Different methods have been proposed, and they generally try to compute the seam for which the discontinuities between the aligned images are less pronounced.

Image blending compensates eventual exposure differences between images, especially on the regions close to the seam. The effect of image blending on a stitched image is illustrated in figure 3.7.

(a) (b)

Figure 3.7: Effect of image blending on a stitched image. (a) Using image blending; (b) not using image blending.

3.5.4 OpenCV Stitching Module

OpenCV has a stitching module, based on the approach proposed by [34]. The functionality is provided by the *Stitcher* class and the processing pipeline is represented in figure 3.8.

Each stage of the pipeline is implemented as a dependency of the Stitcher class. Even though these dependencies are initialized by default in the class constructor, objects with different functionality can be injected through setter methods and thus integrated in the pipeline.

3.6 Classification

Classification is the problem of assigning a given observation to one of multiple classes, using a set of features relative to that new observation. In the field of machine learning, two types of classification are considered:

Supervised classification a set of correctly labelled observations (the ground truth) is available, and is used to build the classifier model.

Unsupervised classification also referred to as clustering, no labelled data is used, and the observations are assigned to different classes (clusters) according to a specific distance metric.

In order to classify the MA candidates as false candidates or true MAs, supervised classification is used, since the dataset used in the development of the MA detection

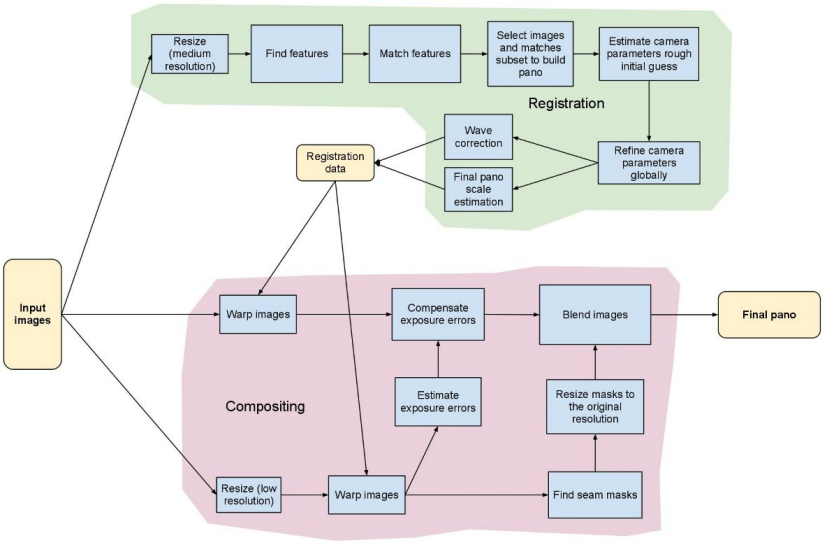


Figure 3.8: OpenCV Stitching Pipeline. From

methodology provides a ground truth. A brief review of two popular classifiers is provided in next sections.

3.6.1 k Nearest Neighbours (kNN)

k-NN is a widely used classifier, which also has applications in regression. For each new observation, this classifier computes a distance metric (for example, Euclidian, Mahalanobis or Hamming) between the feature vectors of the unlabelled observation and the data in the training (labelled) dataset. The k (which is previously defined) nearest neighbours of the new observation, according to the computed distance metric vote the class of the new observation, according to their own classes. As such, for $k = 5$, if three of the nearest neighbours belong to class 1 and the remaining two nearest neighbours belong to class 2, the new observation will be assigned to class 1. An illustration of this process and the effect of the value of k is presented in figure 3.9.

A few considerations should be made regarding some specific aspects of k-NN. Firstly, the k parameter, for a binary classification (two classes) problem should be an odd value, in order to ensure that the voting will never end in a draw. A problem that arises from the voting scheme methodology of k-NN is that when the dataset is unbalanced (the number of elements in a class is much higher than the number of elements in another class), the classification will be biased to the class that dominates the training dataset, since the probability of a nearest neighbour being of the dominant class will be much higher.

Another disadvantage of k-NN is that the training set is directly used to perform the classification and as such, it has to be stored. Furthermore, the computational complexity

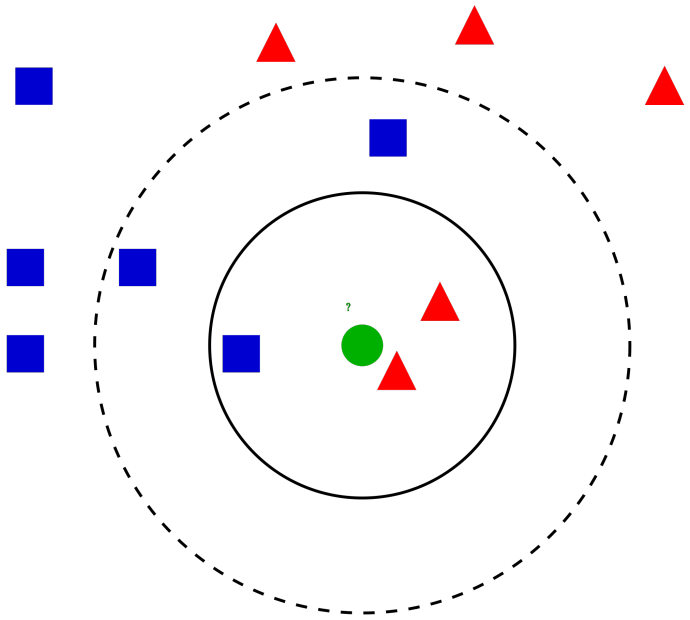


Figure 3.9: k-NN voting procedure. The sample to be classified is represented as a green circle. In case $k = 3$ (contiguous circle), the class will be classified as a *red triangle*, but if $k = 5$ (dashed circle) it will be classified as a *blue square*.

of the classification will vary linearly with the size of this dataset, since each unseen observation will need to be compared with each element in the dataset. This can lead to difficult issues when large training datasets are used.

3.6.2 Support Vector Machines (SVM)

Support Vector Machines is another classifier method, which has the property of computing large margin decision boundaries. These decision boundaries correspond to hyperplanes in the feature space which separate the classes.

In order to compute large margin decision boundaries efficiently, only the support vectors in the data are used. These support vectors are simply the data points that lie closer to the boundary and, as such are the most appropriate to help calculating it - as they are the most difficult samples to classify. Figure 3.10 represents an example of an hyperplane and respective support vectors.

The computation of large margin decision boundaries makes the classifier models generally less prone to overfitting - when a model performs well on the training dataset but fails to generalize its performance to unseen data.

3.6.3 Feature Selection

Feature selection is a process in which a subset of the original set of features is selected. The remaining features will be discarded and only the ones selected will be used for

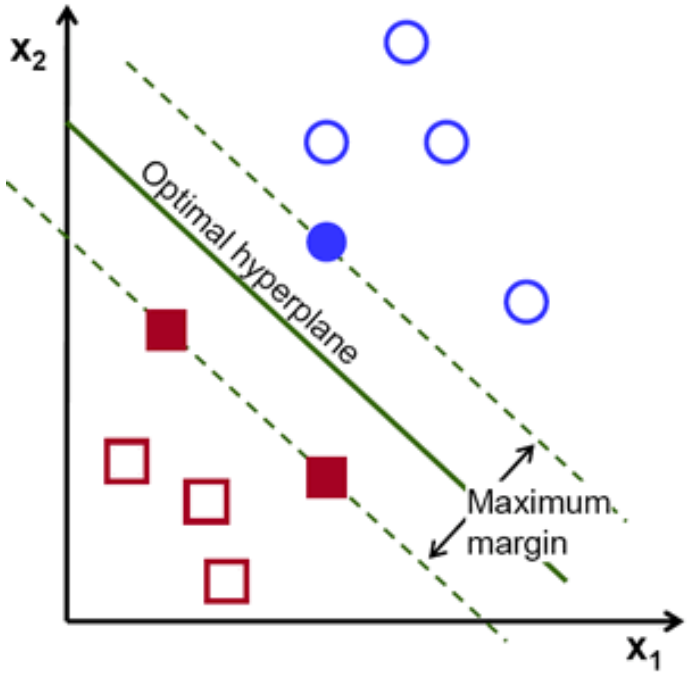


Figure 3.10: Illustration of the SVM computation of a large margin decision boundary. A two dimensional space (two features) is considered. The support vectors for each class are represented as filled circles and squares. From *OpenCV Documentation*.

classification.

The motivation behind feature selection is to reduce the training time, to discard irrelevant/redundant features and improve the generalisation of the classification model. Forward feature selection and backward feature selection are two methods widely used to perform this task.

Forward feature selection is a greedy algorithm, which iteratively chooses the best combination of features, according to some metric - generally, the accuracy - on a cross validation dataset. This method starts by training a model with all the available features individually. Evaluation is then performed and the feature which yields the best results is added to the final subset. The process is repeated, by evaluating the performance of each remaining feature combined with the feature(s) in the final subset, until the performance metric no longer improves.

Similarly to forward feature selection, backward feature selection is a greedy algorithm which iteratively eliminates one feature starting from the original set, until the performance metric no longer improves.

Chapter 4

Exam Guidance

An important objective of this thesis is to develop automatic guidance in order to aid the user during the acquisition of the images. This guidance should instruct the user to rotate the device camera in order to acquire appropriate views of the fundus, which should include the regions around the optic disc.

Guiding the user can be accomplished by tracking the location of retinal structures, while using the history of localizations to indicate how the user should proceed with the acquisition. Candidate structures for being tracked include the optic disc and the macula.

Tracking these structures poses an intricate problem: not only do the algorithms that perform the localization of the retinal structures need to be robust in their estimations, but also need to be very computationally efficient - as they have to run in real time, on mobile devices.

Localization (and segmentation) of the optic disc and macula is one of the main goals of fundoscopy CAD solutions. In general, optic disc localization is generally easier to perform, and as such it will be used.

4.1 Optic Disc Localization

As mentioned, finding the location of the optic disc is a goal of several retinal image processing solutions and given the interest of this task, a variety of methods have been proposed. A widely used method is template matching [35, 36], and due to its simplicity it was also used in this project.

Template matching is a technique used to locate a small subimage (the template) in a larger image. This is accomplished by sliding the template across each point of the original image. For each point, a correlation/matching metric is computed. Example of some metrics implemented in OpenCV are:

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2 \tag{4.1}$$

$$R(x,y) = \frac{\sum_{x',y'}(T(x',y') - I(x+x',y+y'))^2}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}} \tag{4.2}$$

$$R(x,y) = \sum_{x',y'} T(x',y') \cdot I(x+x',y+y') \tag{4.3}$$

$$R(x,y) = \frac{\sum_{x',y'} T(x',y') \cdot I(x+x',y+y')}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}} \tag{4.4}$$

For the metrics 4.1 and 4.2, the best matches are the minimum values and for the metrics 4.3 and 4.4, the best matches are the image maximum values. Experimentally, the metric 4.4 showed more consistent performance and so was chosen.

The template used was a binary image, containing only a white circle in a dark background, as represented in figure 4.1. The background has a padding corresponding to three times the circle radius. This template mimics the intensities of the optic disc and neighbouring regions in the green channel, as illustrated in figure 4.2.

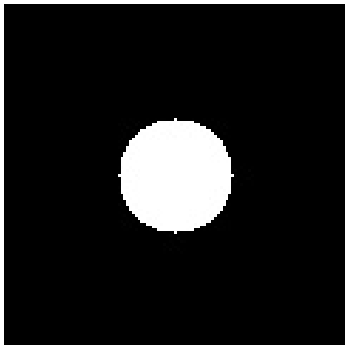


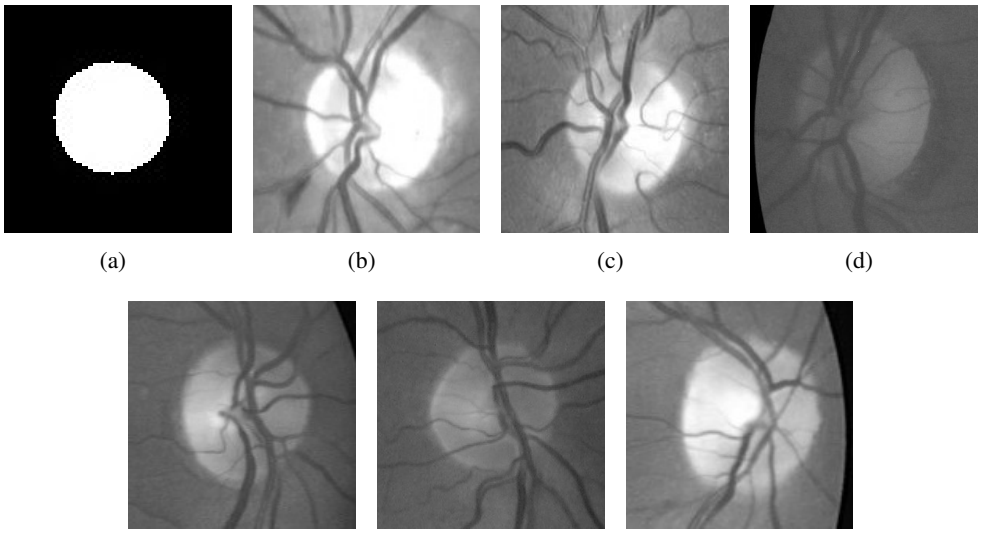
Figure 4.1: Template used for optic disc localization.

In order to account for variation of the optic disc size, three different template sizes are used, maintaining the same padding relation to the circle radius. The disc location is attributed to the point with maximum intensity on the correlation image.

This approach was tested on the DRIVE and ROC dataset (a total of 140 images) with an accuracy of 96.42%. The complete results are shown in table 4.1.

Table 4.1: Results of optic disc location in the DRIVE and ROC datasets.

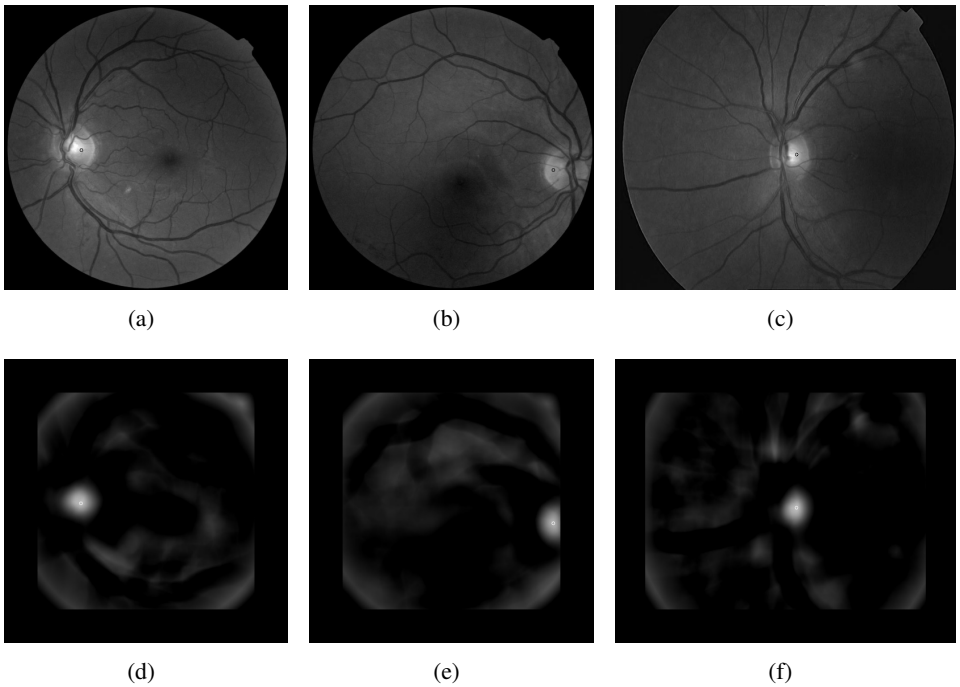
	True Location	**False Location**	**Accuracy**
ROC Training	49	1	98%
ROC Test	46	4	92%
DRIVE Training	20	0	100%
DRIVE Test	20	0	100%
Overall Results	135	5	96.42%



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

(e) (f) (g)

Figure 4.2: Comparison of (a) template (cropped) with (b, c, d, e, f, g) examples of optic discs (green channel images).



(a) (b) (c)

(d) (e) (f)

Figure 4.3: Example of optic disc localization. In (a, b, c) Original images; The optic disc attributed location is marked with a black mark; (d, e, f) Correlation images; The optic disc attributed location is marked with a white mark.

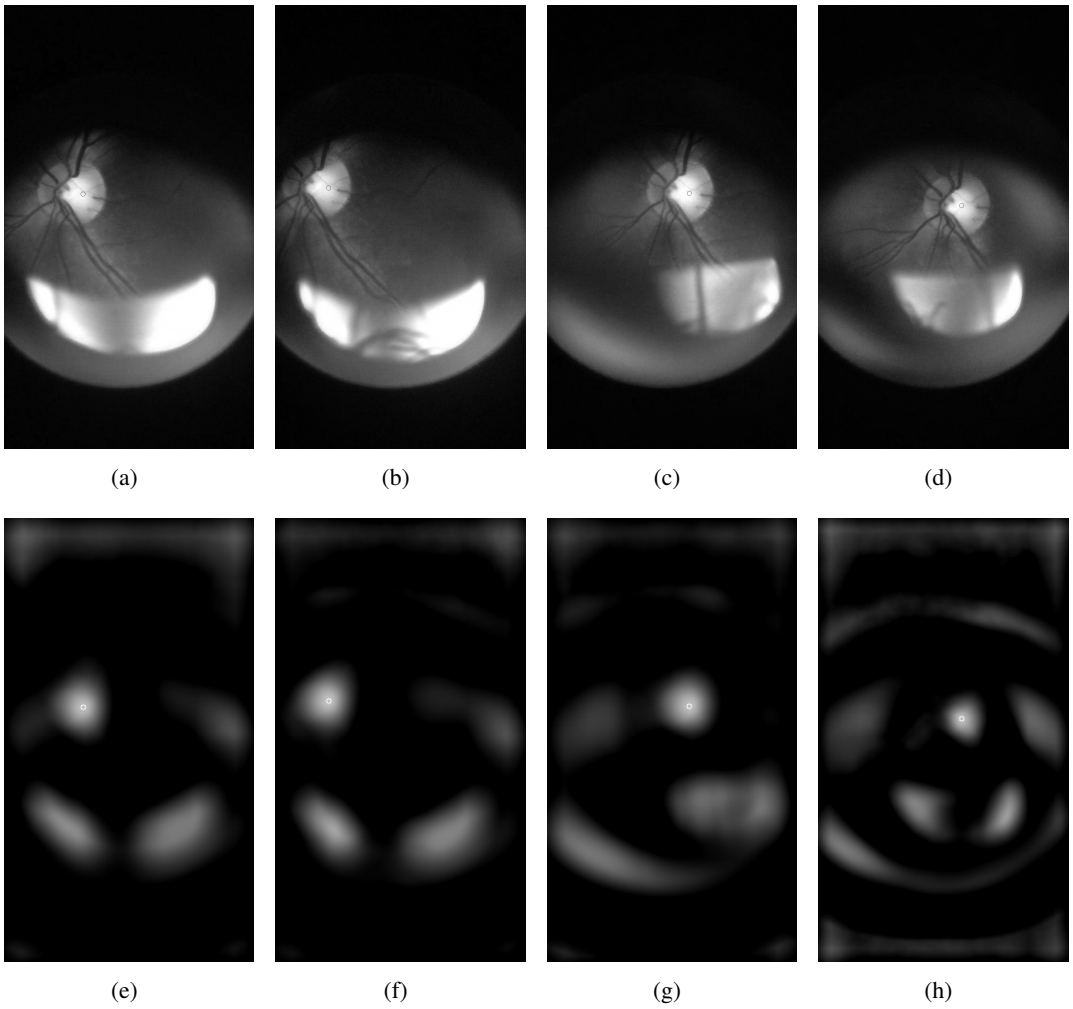


Figure 4.4: Example of optic disc localization in PanOptic acquired images. In (a, b, c) Original images; (d, e, f) Correlation images.

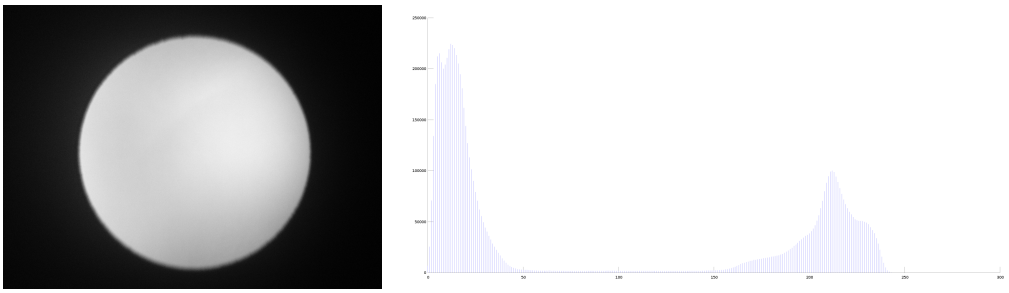
4.2 FOV Calibration

In order to obtain the relative location of the structure being tracked, the available ophthalmoscope FOV should be known. This operation only needs to be performed once for each smartphone, since the mechanical adapter holding the ophthalmoscope will ensure that the FOV will always be the same.

In order to obtain this FOV, the user is instructed to direct the smartphone to a white surface and an image is acquired.

The image is then thresholded, using the threshold computed by *Otsu's* method. The *Otsu's* method automatically computes the threshold of an image from the intensity histogram, iterating over the possible values and choosing the one that minimizes the intra class variability and maximizes the inter class variability [37].

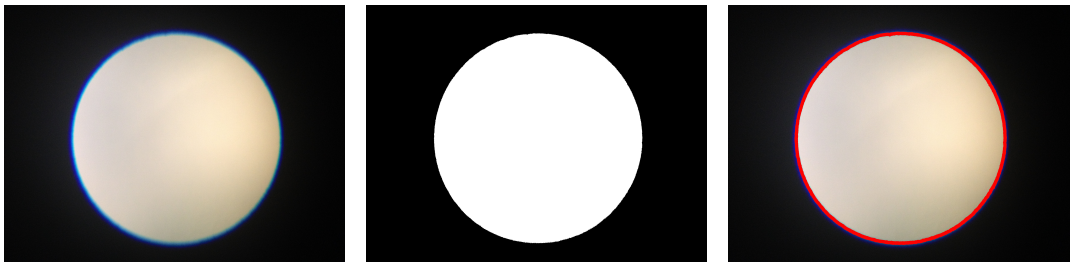
An example of the application of this method is shown in figure 4.5.



(a) (b)

Figure 4.5: Illustration of Otsu's method in the segmentation of the FOV. (a) Grayscale image of the FOV (b) Histogram of the image. The threshold computed by the Otsu's method is 111.

The boundary of the FOV is then saved to be used in the general guidance process (see figure 4.6).



(a) (b) (c)

Figure 4.6: Representation of the different stages for FOV segmentation. (a) Acquired image; (b) Thresholded image; (c) Acquired image, with FOV contours drawn.

4.3 Providing guidance

In order to provide guidance, the FOV segmentation and the optic disc location are used. The FOV is divided into four quadrants and the user is instructed to move the camera so that the optic disc is located in each of those quadrants.

A specific quadrant is only marked as *complete* if the optic disc is located in it at least three times, in order to acquire a sufficient large amount of images. Since the images are acquired continuously, this process does not increase significantly the examination time.

Chapter 5

Retinal Image Stitching

Even though the general image stitching problem is well studied, the specific case of stitching retinal images presents some challenges. Most applications of stitching involve building a landscape or other type of photographic panorama, in which images have well defined interest points, but much of the regions in retinal images are featureless, only presenting gradual intensity variations [38]. The exceptions are the vessels and the optic disc (and eventual lesions).

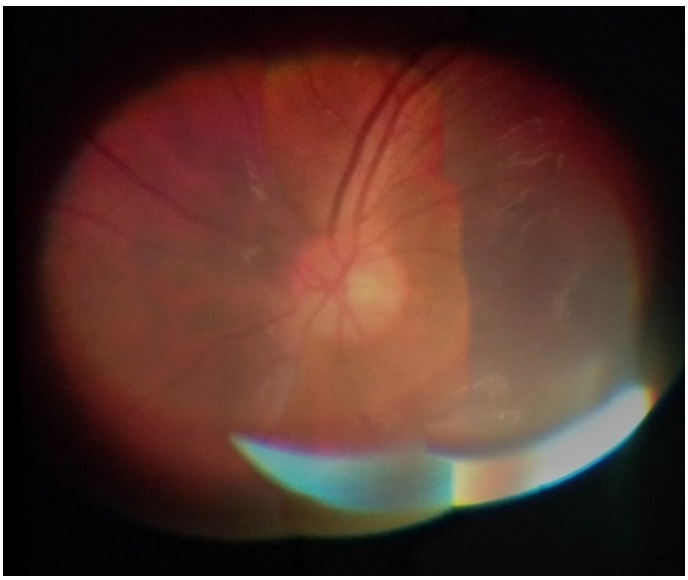
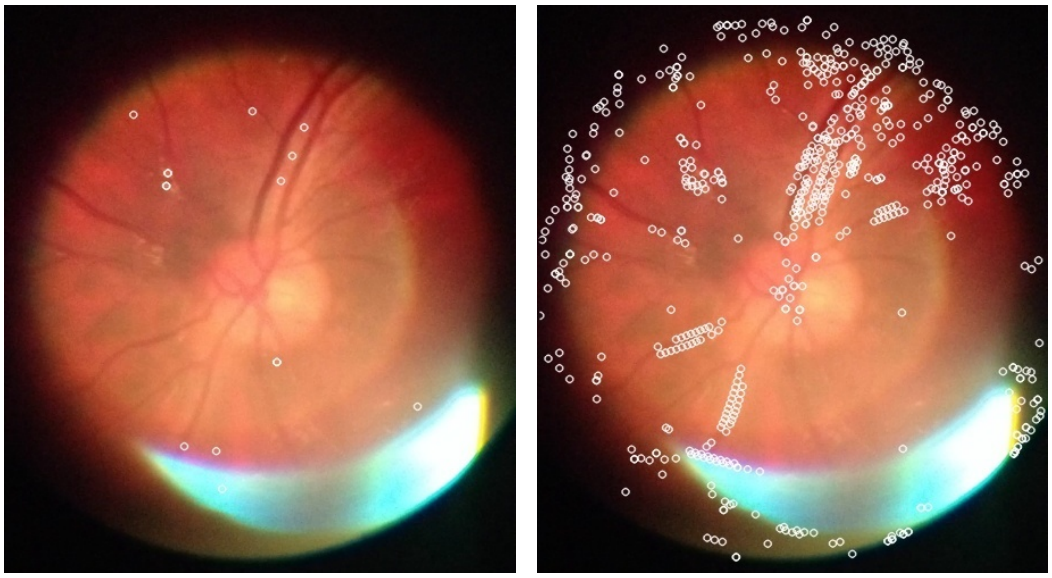


Figure 5.1: Example of retinal image stitching.

The lack of detectable interest points poses a problem for the computation of the alignment (particularly when using feature based alignment), which depends on the robust detection and description of these interest points. In order to address this problem, interest points and respective descriptor vectors are extracted on vessel enhanced images, such as the images used to compute the vessel segmentation in the MA detection pipeline (please refer to section 6.2). Figure 5.2 illustrates this problem.

The stitching functionality was implemented using the OpenCV stitching module, briefly reviewed in section 3.5.4.



(a) (b)

Figure 5.2: Interest points (marked with white circles) extracted from the original image (a) and from a vessel enhanced image (b). The keypoints were extracted using SIFT.

5.1 Vessel Enhancement

As mentioned, vessel enhancement is necessary in order to obtain a robust alignment. As will be reviewed in section 6.2, different methods were developed for vessel segmentation. A similarity between all the methods is that they obtain the final segmentation by thresholding vessel enhanced images. These images can thus be used in order to facilitate the estimation of the alignment. In figure 5.3, examples of these vessel enhanced images are displayed.

Performance of the three vessel segmentation methods was compared, by calculating the number of obtained matches between pairs of overlapping images. The results are shown in table 5.1. The number of existent matches across images is much larger for the Gabor vessel enhancement approach, and so it was chosen.

A common problem with smartphone acquired retinal images is the presence of specular reflections, which result from the reflection of light in the cornea. These reflections may degrade the performance of image processing techniques, particularly the ones used in vessel enhancement. In order to avoid these problems, specular reflections regions were

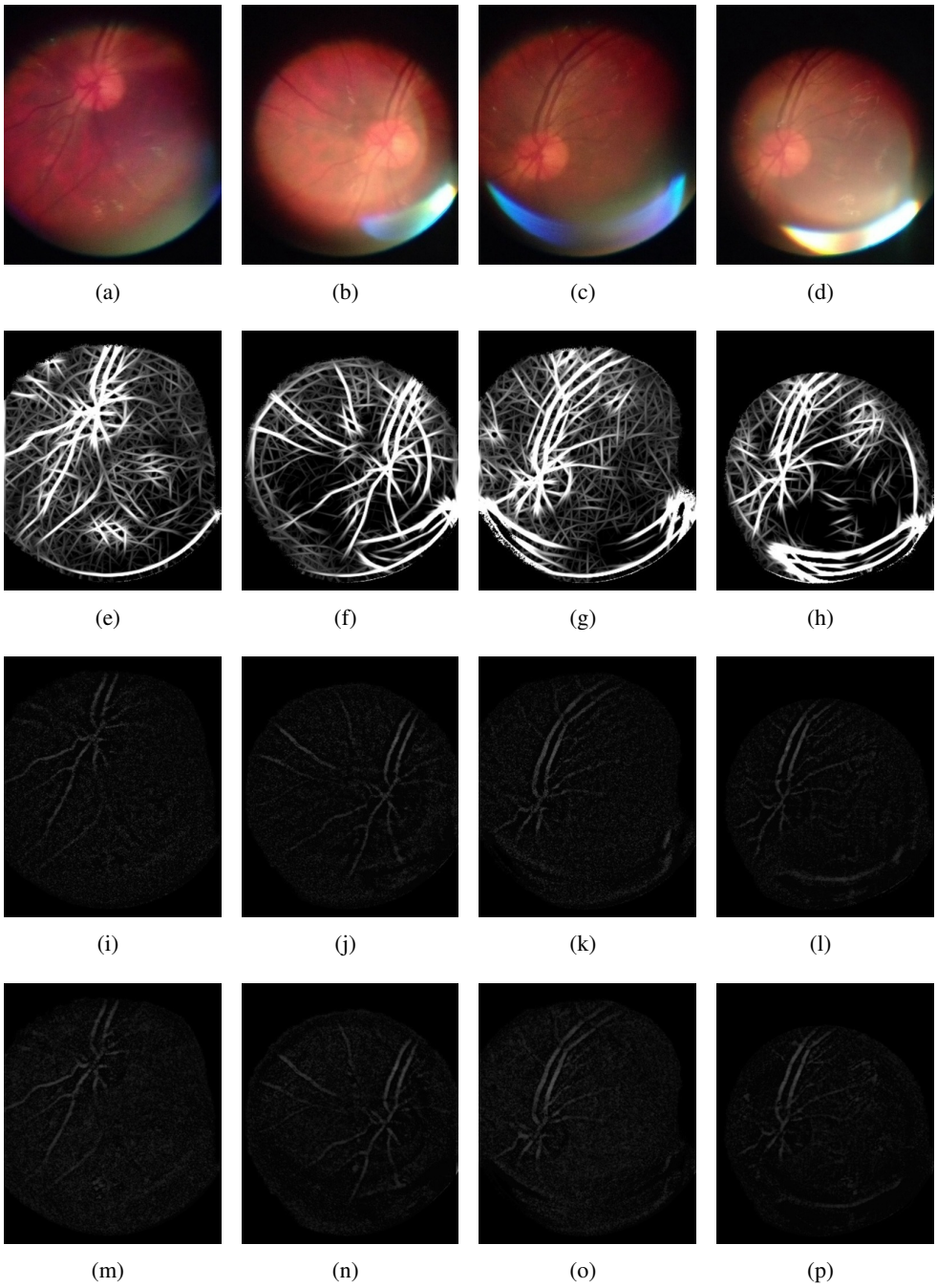
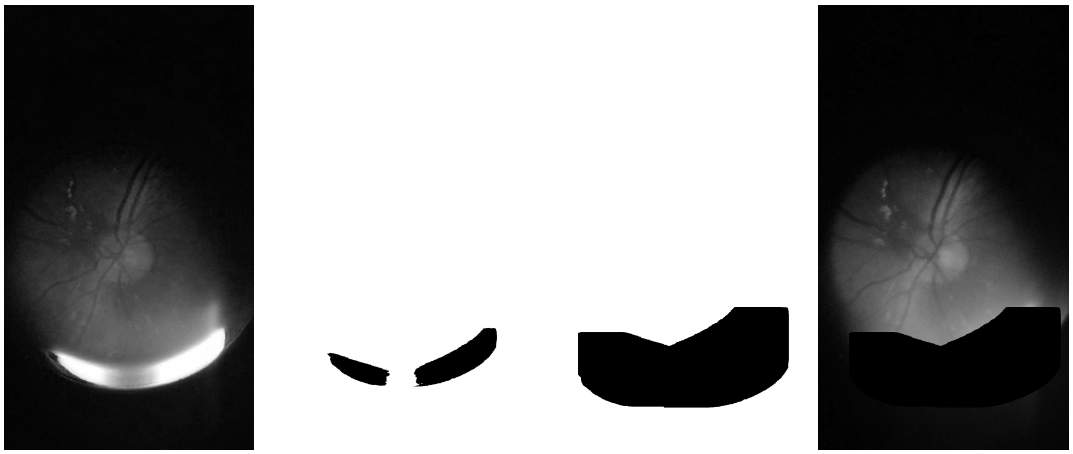


Figure 5.3: (a, b, c, d) Original Images; (e, f, g, h) Gabor vessel enhancement; (i, j, k, l) SCM vessel enhancement; (m, n, o, p) NCM vessel enhancement.

masked, by thresholding the blue channel of the retinal images. The result of the thresholding was eroded in order to obtain a more complete mask. These steps are illustrated in figure 5.4.

Table 5.1: Number of matches obtained using different vessel enhancement methods. The feature descriptor chosen was SIFT and each pair of interest points was classified as a match using the same distance threshold.

	Pair 1 Matches	**Pair 2 Matches**	**Pair 3 Matches**
NCM Enhancement	2	2	1
SCM Enhancement	1	4	2
Gabor Enhancement	32	19	9



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

Figure 5.4: Steps of the specular reflection masking. (a) Original blue channel image; (b) Result of thresholding; (b) Result of mask erosion; (c) Mask applied to the blue channel image.

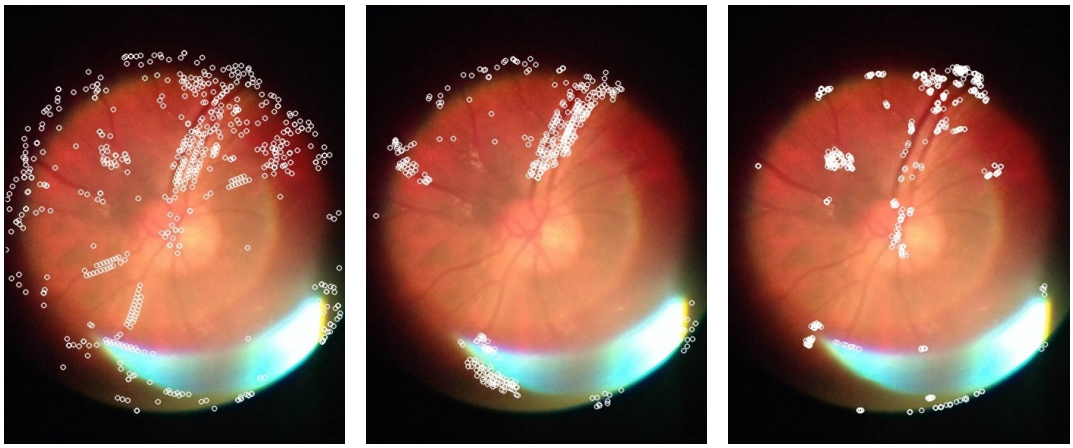
5.2 Feature Based Alignment

As previously reviewed, vessel enhanced images are used for feature based alignment. After choosing the appropriate vessel enhancement algorithm, the next step of this approach is to choose the optimal feature extractor/descriptor method.

SIFT, SURF and ORB were tested, and SIFT consistently outperformed the other methods, in the number of extracted features and consistency of matches. These conclusions are in line with the literature [39, 40]. An illustration of the difference in the number of extracted features is shown in figure 5.5.

The transformation model chosen was the affine transformation. While perspective transforms were tested, the computed alignment was often erroneous, even with high percentages of true matches and the use of RANSAC. This is explained (as reviewed in section 3.5.2.1) by the fact that the probability of finding the correct alignment using RANSAC greatly diminishes with the increase in the number of parameters to be estimated.

Matching of the interest points is performed according to the euclidean distances between the feature vectors. Pairs of points are considered matches if the distance between



(a) (b) (c)

Figure 5.5: (a) SIFT extractor; (b) SURF extractor; (c) ORB Extractor.

their descriptor vectors is inferior to 0.6. Figure 5.6 illustrates the matching between a pair of retinal images.

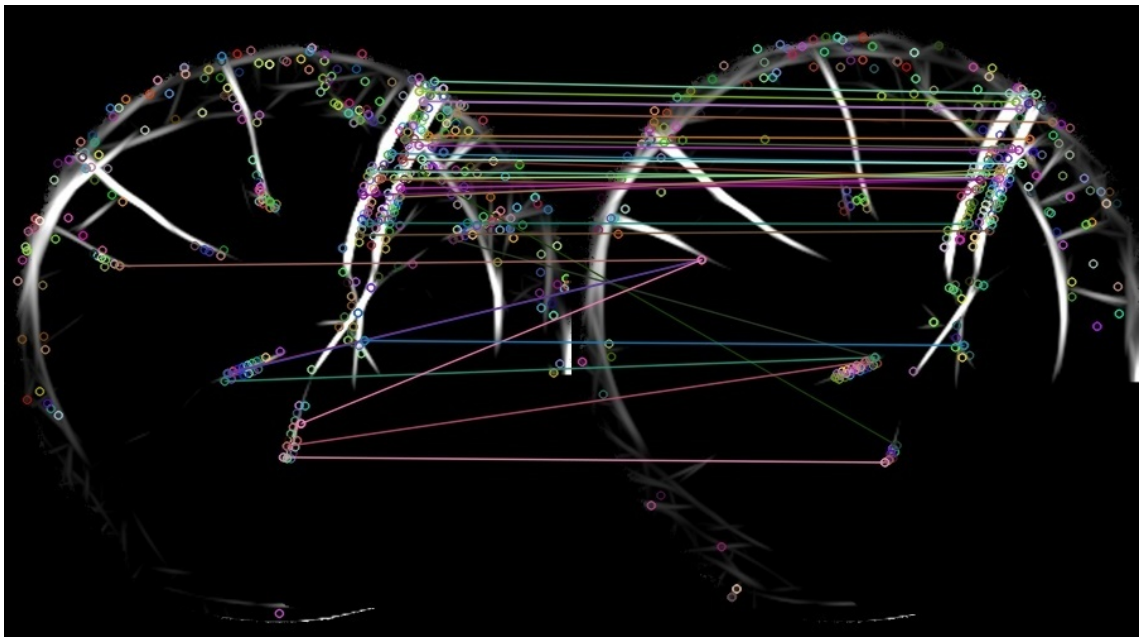
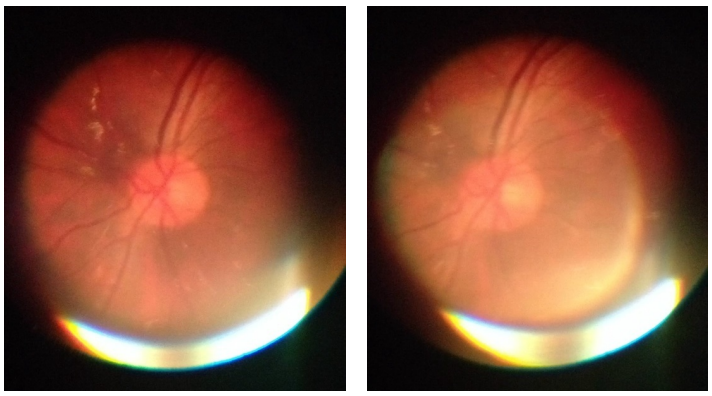


Figure 5.6: Retinal image matching (with Gabor vessel enhanced images). Some false matches can be noted, but using a suitable transformation model, RANSAC is able to estimate the correct alignment.

5.3 Direct Alignment

While the problem of computing the alignment in the original retinal images is more obvious in feature based alignment, direct alignment also suffers from this problem. In

this case, it is not caused by the difficulty in finding interest points, but rather by the significant variations of intensity for the same regions in different images. This problem is illustrated in figure 5.7.



(a) (b)

Figure 5.7: Illustration of the intensity variation of similar retinal regions.

The large variations of intensity will distort the error metric chosen, leading to the wrong calculation of alignments. In contrast, using vessel enhanced images the intensity variations on non vessel regions are removed in both images, as those regions will always have very low intensities. Due to this situation, vessel enhanced images were also used for direct alignment.

As with feature based alignment, several transformation were tested and the similarity transformation (please refer to section 3.5.1.4) was chosen.

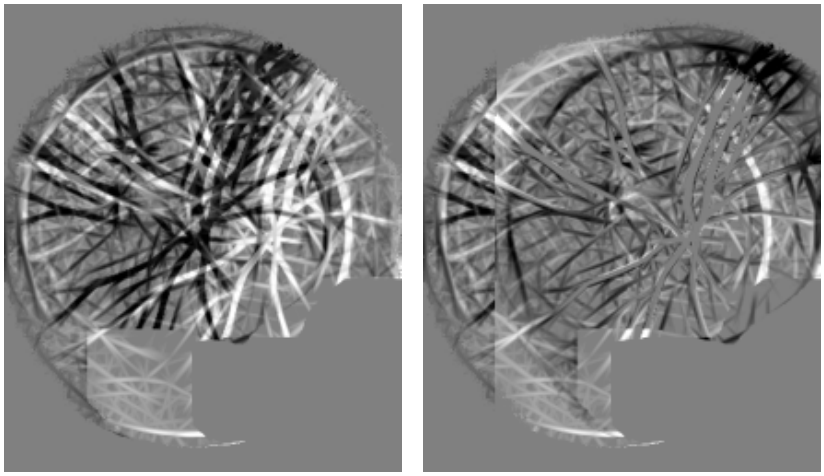
5.3.1 Implementation using OpenCV Stitcher pipeline

While direct alignment presents some interesting advantages over feature based alignment, it is not integrated in the OpenCV stitching module.As such, an implementation of direct alignment was used by adapting a contribution [33] to the development version of OpenCV 3.0.

In order to integrate this functionality in the Stitcher pipeline, the *FeatureMatcher* virtual class was subclassed and customized to perform the direct alignment. Downsampling of the input images is performed for the sake of coping with the relatively high computational cost of direct alignment. The relevant matrix parameters were posteriorly transformed in order to account for this operation. Given a scaling factor s and being $H_s = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix}$ the computed transformation matrix, the final transformation matrix H is:

$$H_f = \begin{bmatrix} a_{00} & a_{01} & s*a_{02} \\ a_{10} & a_{11} & s*a_{12} \end{bmatrix}$$

Due to the ingrained implementation of feature-based alignment in the OpenCV stitching module, the *Stitcher* class includes a *FeaturesFinder* object, which by default uses the ORB feature finder and keypoint descriptor method. Since finding features and computing feature descriptors are costly operations, a custom object which does not perform any image processing operations was set as the feature finder object in the Stitcher pipeline.



(a) (b)

Figure 5.8: Difference between overlapping vessel enhanced images (a) before direct alignment; (b) after estimation of direct alignment.

5.4 Detecting Wrong Alignments

Since the goal of the stitching is to automatically merge acquired images, it may occur that these images may not overlap - for example, the patient might blink or the camera may momentarily lose focus. These cases need to be identified in order to output a high quality stitched image.

In the case of feature based alignment this is accomplished by rejecting geometric transformations computed with a low number of matches. Since the number of matching keypoints is correlated with the similarity of the images, this approach proves robust.

In the case of direct alignment, the SSD (Sum of Squared Differences) of the images is used. As such, the ratio between the SSD of the aligned images with the SSD of the inital images is computed. Only alignments with a ratio lower than a certain threshold were considered.

5.5 Compositing

Composing the stitched image is performed using the default stitching pipeline without any substantial modifications. While OpenCV offers state of the art methods for several of the compositing steps, in practice they offer limited advantages for the specific case of retinal image stitching.

Between the different implemented seam estimation methods, Graph Cut was chosen. Image blending was not used, since experimentally it often led to blurring of the stitched image.

5.6 Feature based alignment vs Direct Alignment

Stitching results for both alignment methods are shown in figure 5.9. The accuracy is generally higher for direct alignment, as is evident in most examples: while feature based alignment computed a close estimation of the correct alignment, some discontinuities are noticeable, and these do not occur for direct alignment. However, direct alignment often fails to compute the transformation where only a small overlaps exists.

Due to the intrinsic characteristics of the developed acquisition methodology (periodic image acquisition, small motions between acquisitions), this disadvantage of direct alignment is not very relevant. Furthermore, besides the worse accuracy of feature based alignment methodology, the SIFT algorithm is protected by intellectual property rights and its computational performance is also worse.

For these reasons, the developed direct alignment algorithm was chosen to be implemented in the Android application.

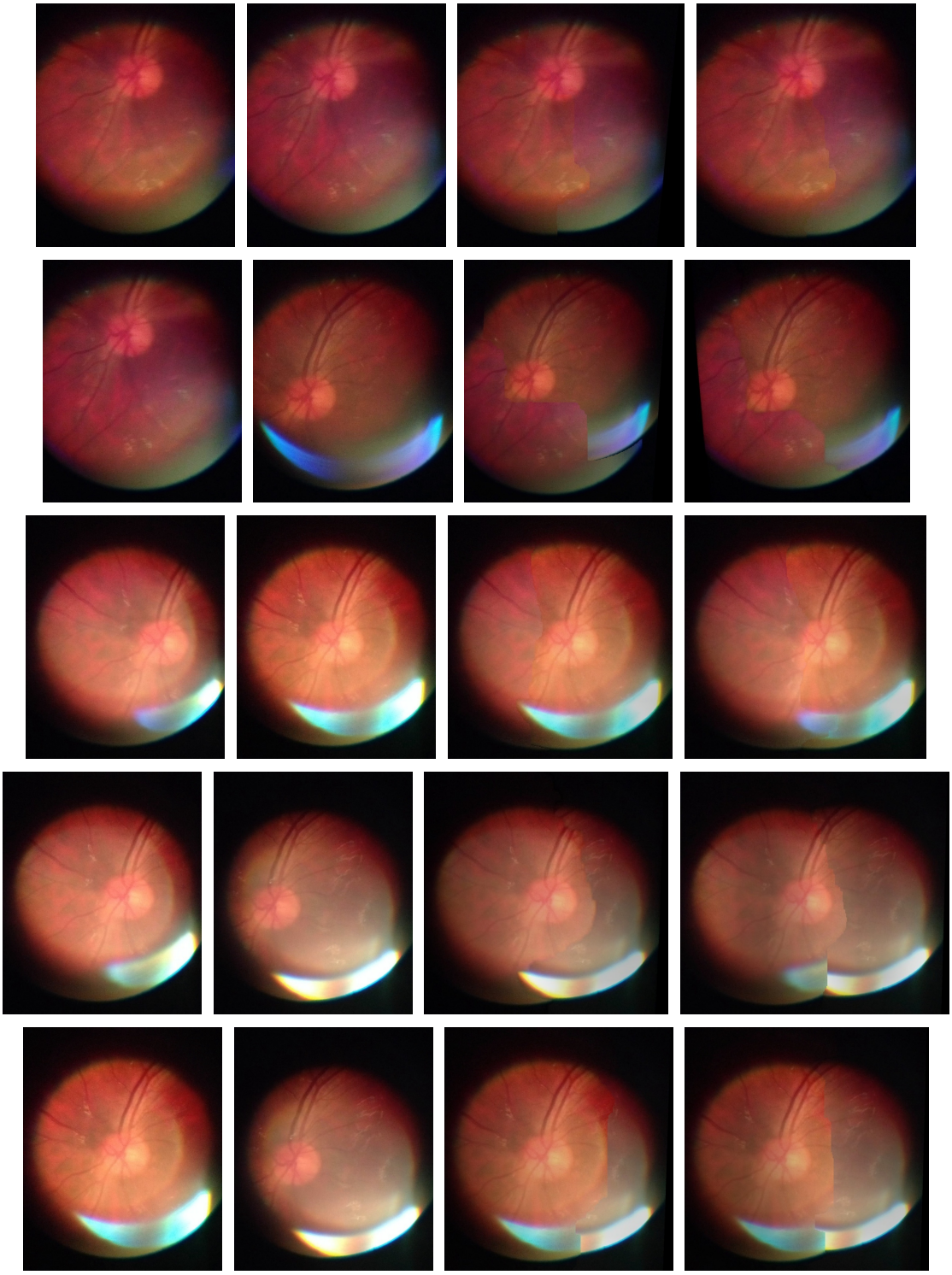


Figure 5.9: Results of stitching. The first and seconds columns are the input images; the third column has the results of the feature based alignment approach and the fourth column has the results of the direct alignment approach. Image blending was not used.

Chapter 6

Microaneurysm Detection

Automatic microaneurysms detection in retinal images is performed using a processing pipeline similar to other state of the art methods, as presented in section 2.4 and illustrated in figure 6.1. On a general level, four different elements can be considered as part of this pipeline: the input retinal image, a candidate extraction step, a classification of the extracted candidates and finally the output, corresponding to the positions of the detected MAs.

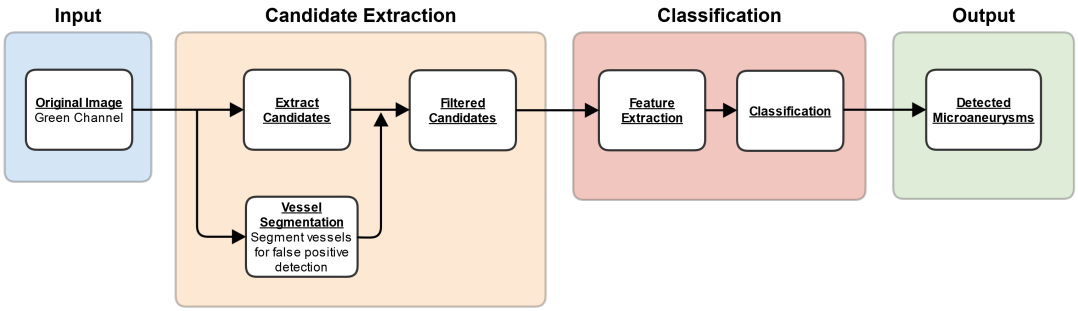


Figure 6.1: Microaneurysm Detection Pipeline.

From the original input image, only the green channel of the RGB color space is used for the main processing steps. The reason for this decision is the superior contrast of this channel, in opposition to the significant oversaturation of the red channel and undersaturation of the blue channel. These conditions are illustrated in annex A.1.

In the candidate extraction step, the green channel of the input image is processed using appropriate image analysis methods. The result of this step is a set of microaneurysm candidates, each composed of a connected set of points on the original image. A common problem that occurs in this step is the detection of several false candidates in the retinal vasculature. In order to address this issue, a vasculature map is obtained and any candidate located on a vessel is discarded.

Classification of the candidate is the final processing step of the MA detection pipeline. In order to perform classification, a set of features is computed for each candidate. The result of the classification will be the localization of the microaneurysms, which will constitute the output of the entire pipeline.

6.1 Candidate Extraction

In retinal images, microaneurysms appear as small (3-10 pixels radius), circular and hypointense (in the green channel) regions. These characteristics are used by candidate extractors to detect their location.

Two different methods that perform extraction of candidates were used: a method based on the detection of circles and a method based on their intensity distribution. Results relative to both methods were obtained for the ROC training dataset, comprising 50 retinal images with and without microaneurysms.

6.1.1 Circles Extractor

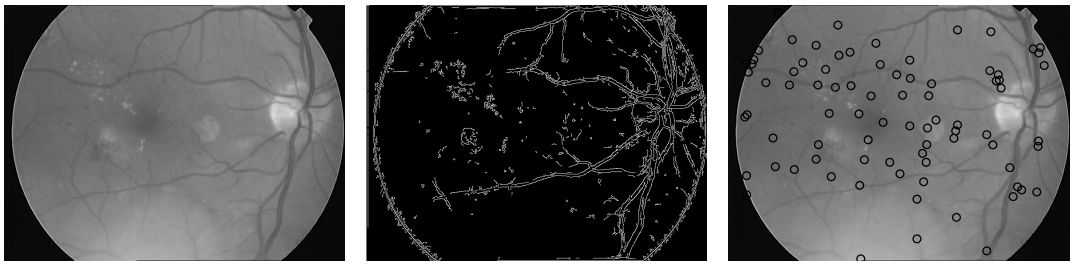
Based on the approach proposed in [41], microaneurysms candidates are extracted by identifying their circular shape. This is performed by using a feature extraction technique named Circle Hough transform (CHT).

In order to use the CHT, an edge detector (in this case, the Canny Edge detector [42]) is employed as a pre-processing step, so that the edges of the original image are obtained. The CHT technique uses a three dimensional accumulator matrix, where each dimension represents one of the three parameters of a circle (x position, y position and radius). All the elements of the accumulator matrix are initially set to zero, being incremented each time a specific edge pixel belongs to the parametrized circle. Finally, the accumulator matrix is thresholded and only elements (corresponding to specific circle parameters) with an accumulator value greater than the threshold are considered as candidates.

In order to apply CHT to the problem of finding microaneurysms, the radius of the circles are limited both inferiorly and superiorly, by 3 and 10 pixels, respectively. The advantage of using CHT is that even if the edges are not correctly detected, the corresponding circle can still be obtained, only with a lower accumulator value. The steps of this method are detailed in figure 6.2.

6.1.2 Gaussian Correlation based extractor

As previously mentioned, microaneurysms present specific characteristics, related to their size, shape and intensity distribution. A possible approach to use in their detection, harnessing the known information about their attributes, is to perform template matching



(a) (b) (c)

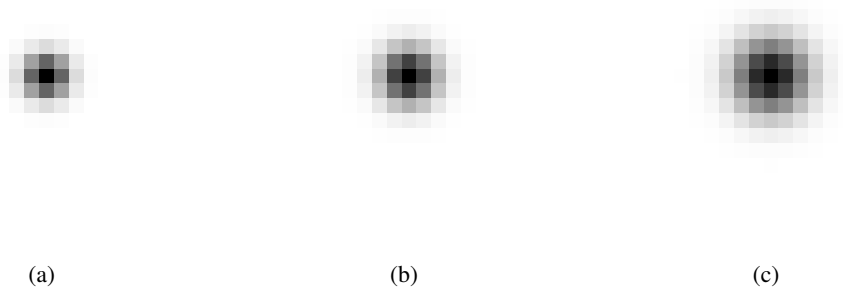
Figure 6.2: Steps of the CHT based candidate extractor. (a) Input image (green channel); (b) Edges obtained by the Canny edge detector; (c) Input image with the extracted candidates marked with black circles.

with an inverted 2D Gaussian kernel, as first proposed by *Zhang 2010* [43]. The 2D inverted Gaussian kernel is modelled by the following function (the mean is considered 0):

$$F(x,y) = -e^{-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)}$$

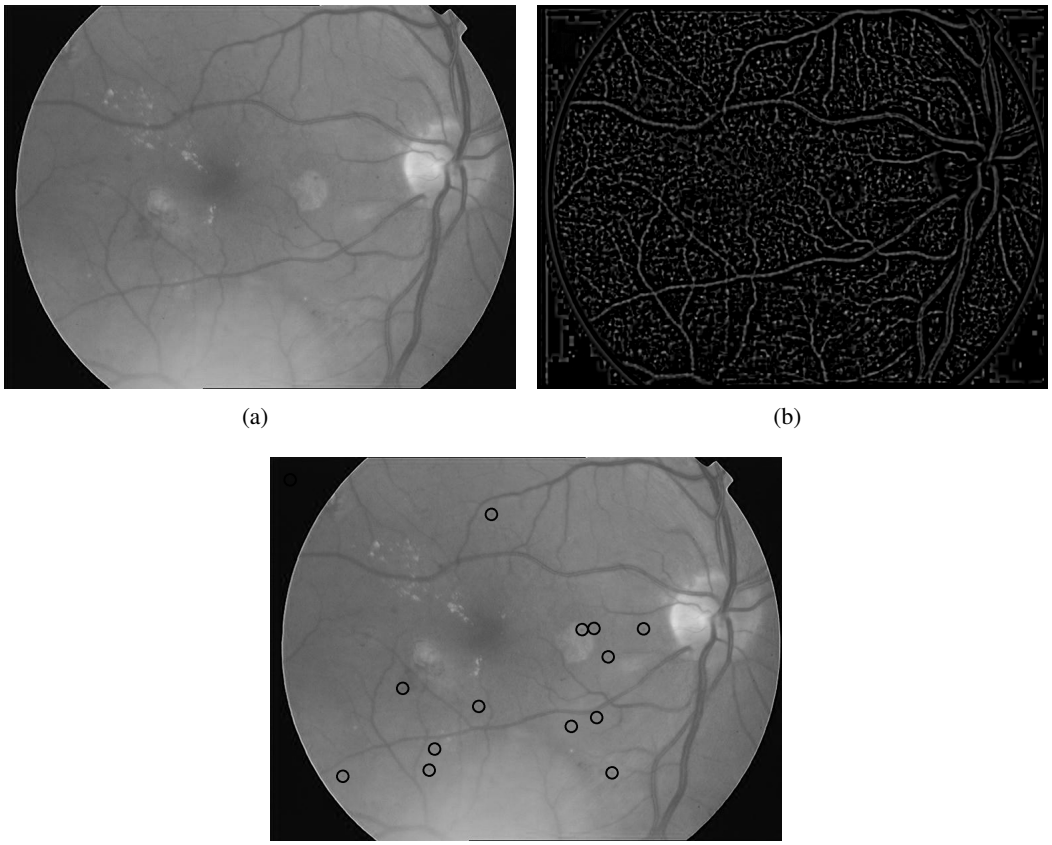
This Gaussian kernel not only emulates the distribution of intensities (lower for the microaneurysm, higher for the background, generally following a normal distribution), but also its shape (by using $\sigma_x = \sigma_y$, the kernel has a circular shape) and size (also defined by the σ value).

Since microaneurysms present vastly varying sizes, five different kernels are used, with different σ values. In figure 6.3, the effect of different sigmas in the Gaussian kernel is illustrated. The result of the correlation for each sigma is thresholded, and the results for the different kernels combined. Each connected component that results from the thresholding operation will constitute a MA candidate. The steps of this algorithm are represented in figure 6.4.



(a) (b) (c)

Figure 6.3: Gaussian kernel variation with *sigma*



(a) (b)

(c)

Figure 6.4: Steps of the Gaussian Correlation based extractor. (a) Input image; (b) Result of correlation with inverted Gaussian kernel; (c) Candidates resultant from thresholding the correlation image, marked with black circles on the input image.

A variety of parameters were tested, such as different sigmas and template sizes. Experimentally, the results obtained showed significant variance, being the final optimal template size 21 and the optimal set of sigmas 1.3, 1.4, 1.5, 1.6 and 1.7.

Besides the sigmas and the template size, varying the threshold value will also render different results. The specific variation of this parameter was used to compute different pairs of sensitivity and number of false positives.

6.1.3 Results

Evaluation of both candidate extractors was performed using the ROC training dataset, which consists of 50 images with a total of 336 microaneurysms. The results reported in this section were obtained using the same parameters for the entire training dataset of ROC, without any modification to account for the different image sizes - only a preliminary resizing step was executed for both detector methods, in order to normalize the width of all the images. The trade-off between sensitivity and number of false positives is

achieved singly by variation of the threshold value on the last step of both methods. The sensitivity value is given on a per lesion basis - a true positive is only considered when the located candidate lies inside the radius of a true microaneurysm. The presented results already include the removal of false positives located on vessels, using the output of the vessel segmentation methodology detailed in section 6.2.2. In figure 6.5, examples of the result of this candidate extraction step are portrayed.

The FROC curve is a plot of the sensitivity (also referred to as the TPR - True Positive Rate) against the number of false positives (in this case, it is the mean number of false positives per image) and this plot portrays the performance of a given detector method. Sensitivity (or TPR) is defined as:

$$TPR = \frac{TP}{TP + FN}$$

The FROC curve for both methods is represented in figure 6.6. The results evidentiate that the GC detector is clearly superior to the CHT detector, for any number of false positives considered. In table 6.1, the two implemented approaches are compared to the similar method developed by Zhang et al [43] and the Math morphology method [17], for four different values of false positives. The results are relative to the same ROC training dataset, including a total of 336 true microaneurysms. The GC detector outperforms the remaining methods for all values of false positives. In table 6.2 the developed implementations are further compared with other state of the art methods. All reported results for these methods are provided for the implementations by [44], with the exception of the *Adal 2010* result [45]. Once again, the GC detector outperforms most approaches in terms of sensitivity, only being surpassed by the Adal et al and Lazar et al implementations, both presenting much higher values of false positives.

Table 6.1: Performance comparison for different number of average FPs per image.

FPs / image	2	4	8
CHT based	3.3%	4.6%	6.3%
GC based	14.9%	23.2%	30.9%
Zhang 2010 [43]	12.6%	21.0%	24.3%
Math Morph [17]	5.5%	10.4%	13.6%

A serious difficulty faced when using the ROC training dataset is the presence of a large fraction of true microaneurysms that exhibit very limited contrast with the background. Figure 6.7 illustrates this issue, which is also reported in related work. Indeed, in [47] two of the co-authors evaluated the images of the ROC training dataset, considering as visible only 196 out of the 336 ground truth microaneurysms. These considerations are further substantiated by the variation in performance of the same methods on different datasets - in *Zhang 2010* [43], for 2 false positives per image, the sensitivity (after the

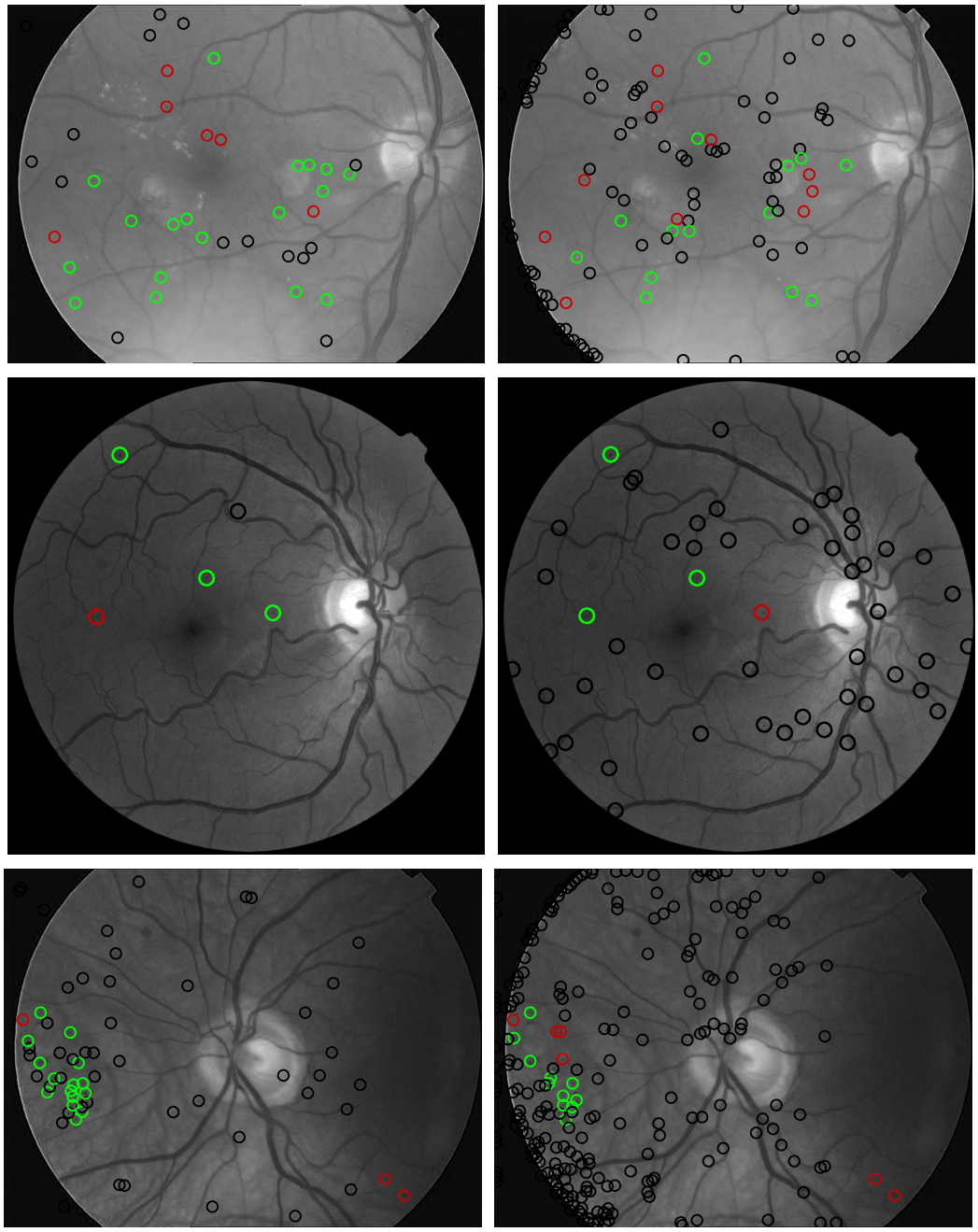


Figure 6.5: Examples of the candidate extraction step results. The left column is relative to the GC candidate extractor and the right column is relative to the CHT candidate extractor. All the green and black circles correspond to the candidates locations obtained by these methods. Green circles mark true microaneurysms, black circles mark false positives and red circles mark not detected true microaneurysms.

classification step) for the ROC training dataset is 26.7%, for the the ROC test dataset is 39.4% and for the DIARETDB1 [48] test dataset is 88.6%.

This problem may result from the high sensitive nature of the ROC training dataset annotation methodology: four experts annotated the images and a microaneurysm was

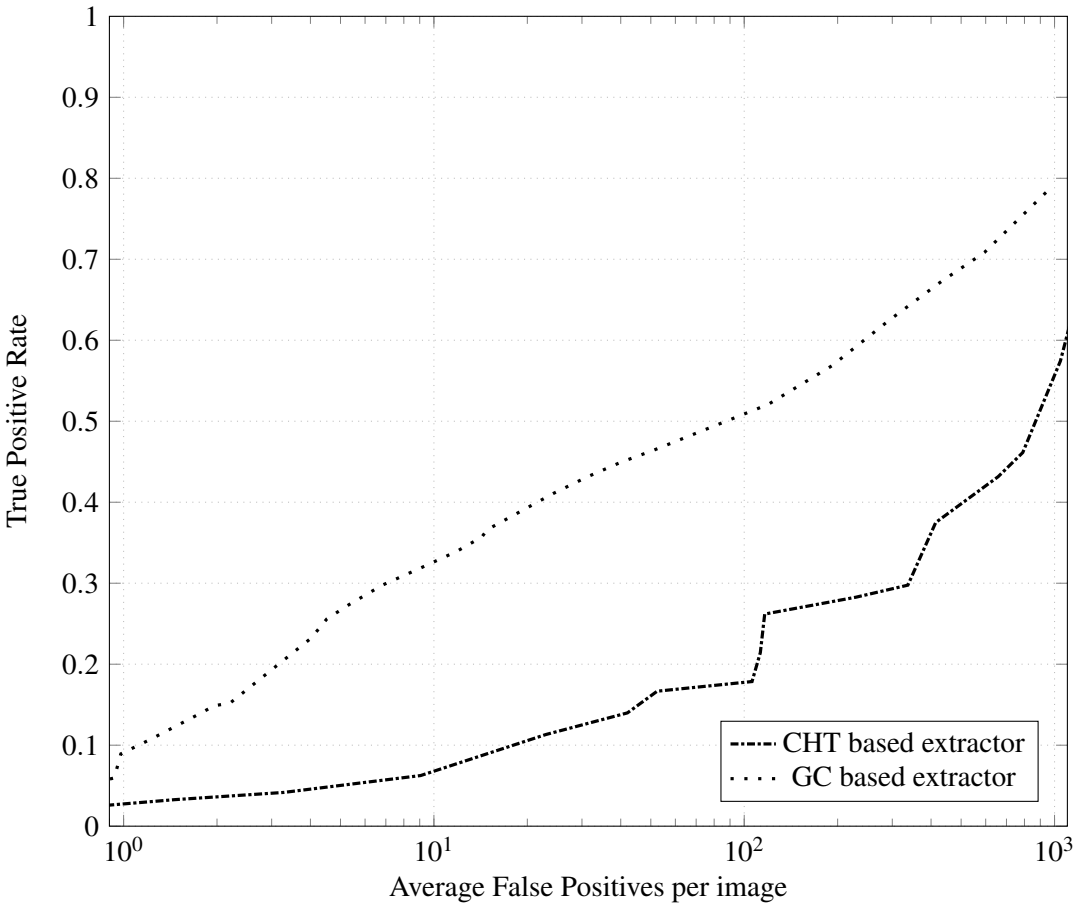


Figure 6.6: FROC plot of the performance of the two implemented candidate extractor methods.

Table 6.2: Comparison of the developed candidate extractor method results with other state of the art methods.

Method	**Sensitivity**	**FPs / Im**
CHT based	16.67%	52.42
GC based	41.07%	24.22
Spencer et al [18]	12%	20.3
Abdelazeem [41]	28%	505.85
Walter et al [46]	36%	154.42
Zhang et al [43]	33%	328.3
Lazar et al [12]	48%	73.94
Adal et al [45]	44.64%	35.20

registered even if it was marked by a single expert. In the test dataset, the ground truth was obtained by using the annotations of three experts, and a microaneurysm was only marked if a consensus existed by at least two of the experts.

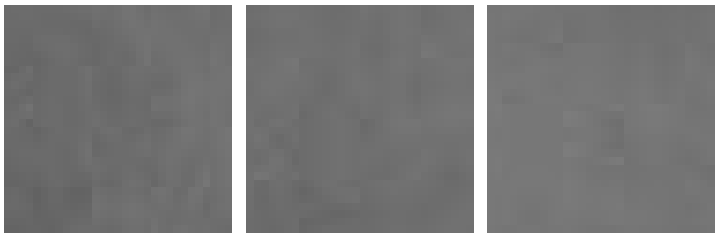


Figure 6.7: Examples of invisible microaneurysms that are included in the ground truth. Each of the images corresponds to a 30×30 window centered in the MA location. Compression artefacts can be detected in all images.

6.2 Vessel Segmentation

The most common location of false positives on the candidate extraction step is on the retinal vessels. Retinal vessels are elongated, connected regions of pixels with low intensities, and as such they are significantly different from MAs. However, since most candidate extractor methods are sensitive to low intensities on edges, it is common to obtain a significant number of false positive candidates on these regions. It should also be noted that even though microaneurysms correspond to enlargements of small vessels, the size of these vessels makes them undetectable on conventional imaging of the fundus, and so removal of candidates in (correctly identified) vessels does not remove true microaneurysms, but removes a significant portion of false positives that may hinder the global performance of the method (particularly the specificity).

A method for vessel segmentation is thus necessary for enabling the removal of these false positive candidates. Significant research in the field of Computer Aided Diagnosis has been focused on the segmentation of retinal vessels, and state of the art methods present accuracy values close to human experts. However, these methods are generally computationally expensive, and also often require a training step.

As the main goal is to extract false positive MA candidates and not to obtain a clinical grade vessel segmentation, a simpler approach should be taken, focused on fast execution and low number of false positives, in order to avoid erroneously discarding candidates that are not truly located in vessels.

Three different methods were devised, based on grayscale morphological operations and Gabor filters. Validation of these methods was performed on the public DRIVE dataset, reviewed in section 3.3.1.

6.2.1 Non Corrected Morphological (NCM) Vessel Segmentation

Since the properties of the vessels in fundoscopy images - low pixel intensities, connected and elongated regions - are well known, they can be used to perform the vessel segmentation.

One possibility is using grayscale morphological operations. Basic operations such as dilation and erosion use a defined structuring element (SE) to transform the original image pixels according to the intensity values of their neighbours. Erosion will attribute to each pixel the lowest value of any of their neighbours within reach of the structuring element and dilation will attribute the maximum value of the neighbours.

Opening and closing are higher order morphological operations, which use erosions and dilations. Opening will perform a dilation over the result of an erosion and closing will perform the inverse (erosion over the result of a dilation). In this method a *bottom hat* procedure was used to enhance retinal vessels. This procedure corresponds to subtracting the result of the original image closing operation by the original image itself.

Bottom hat allows the enhancement of dark structures that are both smaller than the SE and darker than the neighbours and, as such, enhances the vessels. In this method, the result of that *bottom hat* is thresholded according to a fixed value.

As the thresholded image will still include other small structures (for instance, MAs may be included, as they are smaller than the SE and darker than their neighbours), a subsequent step will remove all small connected components. Since vessels are long connected structures, they will not be discarded.

The first and last operation to be performed in this method is a resizing of the original image - initially to a smaller size and in the end back to the original dimensions. The motivation for this is to improve the running time and also to make the method parameters generalizable to different input image sizes. An example of the results for the different steps of the method is presented in figure 6.8.

6.2.2 Shade Corrected Morphological (SCM) Vessel Segmentation

The first step of this method is shade correction. In this step, the original image is filtered with a median filter and then the same original image is subtracted from that result. The median filtering will allow to obtain an approximation of the background of the fundus image (a relatively large kernel size is used). By subtracting the image from the background, contrast between the vessels and the background will be enhanced (as background intensities will be close to zero), and vessels will be hyperintense (bright) structures.

Similarly to the NCM method, a morphological operation will also be used. Instead of *bottom hat*, *top hat* will be employed, which consists of subtracting the image by a opened version of itself. Instead of enhancing darker structures, this will enhance the ones that are brighter (which include the vessels).

Similar steps of thresholding, removal of small connected components and resizing operations were also implemented, with the same goals as in the NCM method. Results for the different steps are represented in figure 6.9.

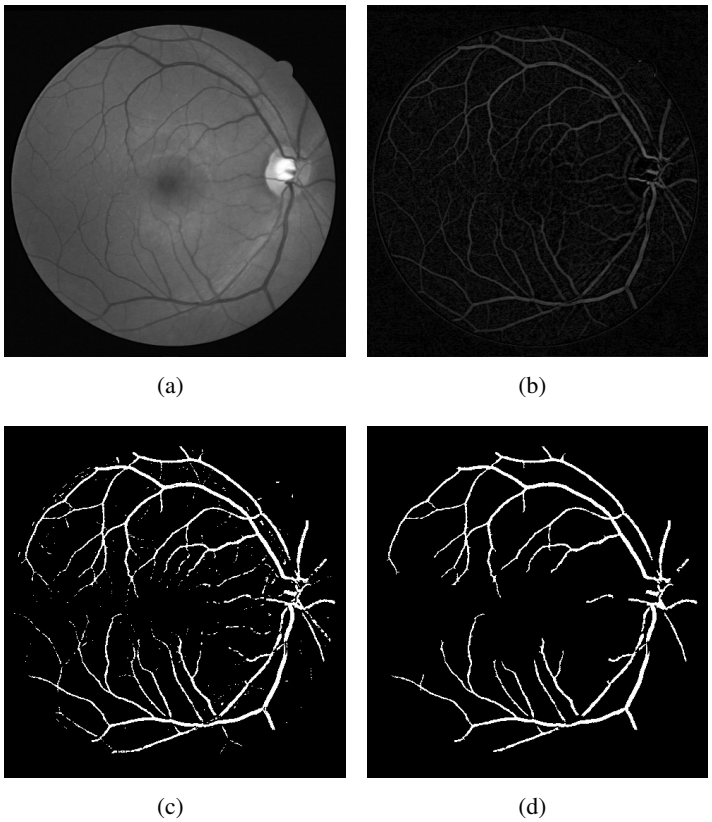


Figure 6.8: Steps of the NCM method. (a) Resized original image; (b) After *Bottom Hat* transform; (c) After thresholding; (d) After small components removal.

6.2.3 Gabor filter based segmentation

Gabor filters are a type of linear filters that are commonly used for feature extraction and edge detection. They are defined as the multiplication of a Gaussian envelope by a complex sinusoid:

$$g(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} e^{j\omega_0 x}$$

A total of twelve different orientations (ω_0) of the Gabor kernel are used, and the respective filter responses are combined. The ω_0 values used are in the interval 0 to π. Examples of different Gabor kernel orientations are shown in figure 6.10. The different kernel orientations are used due to the fact that the filter response will depend on the orientation of the vessels, and in order to obtain the response for all the directions, the entire range of kernel orientations must be used. An example of the response for different kernel orientations is displayed in figure 6.11.

After the results of the successive filter responses are combined, thresholding is performed and the final vessel segmentation is obtained. The various steps of this algorithm are shown in figure 6.12.

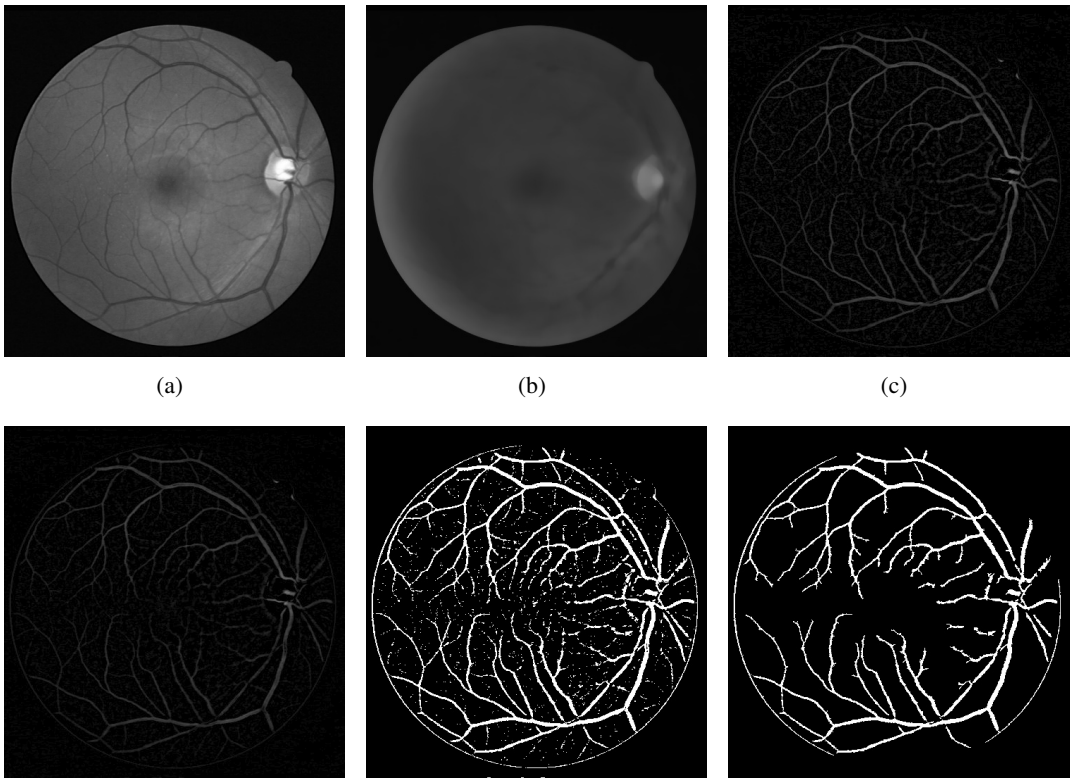


Figure 6.9: Steps of the SCM method. (a) Resized original image; (b) Background estimation by median filtering; (c) After Shade Correction; (d) After *Top Hat* transform; (e) After thresholding; (f) After small components removal.

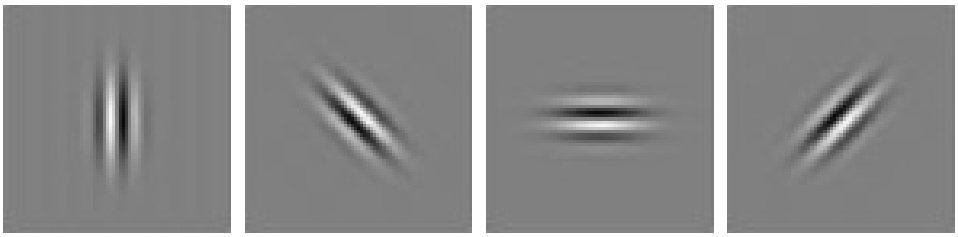
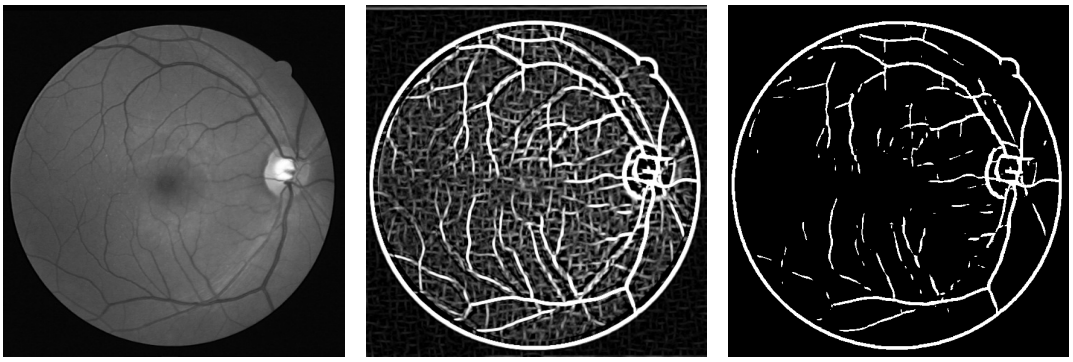


Figure 6.10: Different Gabor kernel orientations.

Figure 6.11: Responses for different Gabor kernel orientations.



(a) (b) (c)

Figure 6.12: Steps of the Gabor vessel segmentation method. (a) Original Image; (b) Result of the combination of the responses to the different kernel orientations; (c) Result after thresholding.

6.2.4 Evaluation

As mentioned before, evaluation of both methods was performed on the DRIVE dataset introduced in section 3.3.1.

Evaluation was performed pixelwise, which means that it took into account if the vessel/non vessel classification was correct for each pixel. Please refer to figure 6.13 for an example of the evaluation results for a single image.

By varying the threshold values on both methods, different levels of segmentation were obtained and used to compute a ROC (Receiver Operating Characteristic) curve. The ROC curve is a plot of the sensitivity (also referred to as the TPR - True Positive Rate) against the false positive rate (FPR), that portrays the performance of a given classifier method.

As previously explained, TPR is defined as

$$TPR = \frac{TP}{TP+FN}$$

and the FPR is defined as

$$FPR = \frac{FP}{FP+TN}$$

where TP is the number of true positives, FP the number of false positives, TN the number of true negatives and FN the number of false negatives.

ROC curves for the developed methods are represented in figure 6.14.The results show that for every FPR value, the TPR is always greater for the SCM method. This is especially evident for low values of FPR (< 0.1), and that is arguably the most relevant region of the curve - a large FPR value is unacceptable for the purpose of false candidates removal, as it may lead to exclusion of true microaneurysms. In general a low FPR is favored over high sensitivity (TPR), and thus a conservative threshold shall be used for

the integration in the global MA detection system.

Accuracy and area under the (ROC) curve are two other relevant metrics for the characterization of a classifier system. A comparison of the results of NCM, SCM and Gabor with state of the art methods and a second human rater segmentation, on the DRIVE dataset, are represented in table 6.3. Once again, results show that the SCM method is superior to the NCM and Gabor methods, for the two metrics. The results for area under the curve are inferior to most existent approaches, which can be explained by the relatively bad performance of the developed methods for large values of FPR. Accuracy values for the SCM are similar to the other state of the art methods.

All the presented values are relative to the performance of the developed and state of the art methods in the test images of the DRIVE dataset. As recommended, pixels outside the image FOV were not considered for the evaluation (see figure 6.13), as their detection as true negatives is trivial.

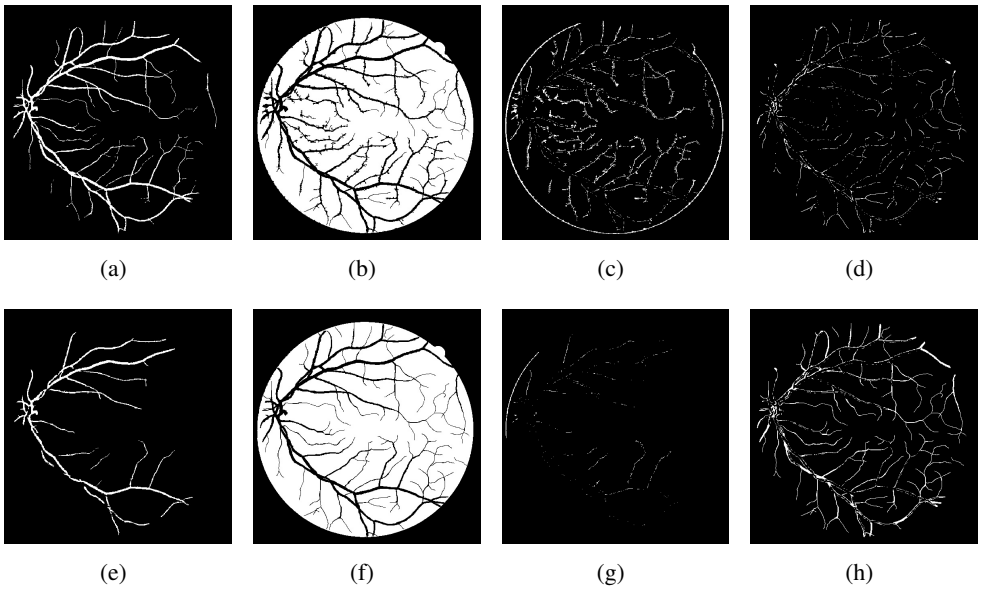
Table 6.3: Comparison of performance metrics of various methods on the DRIVE dataset. Accuracy values were obtained for the optimal threshold values. Performance measurements for the remaining methods were obtained from [1].

Method	**Accuracy**	**Area under ROC**
NCM	0.8946	0.7898
SCM	0.9231	0.8701
Gabor	0.8477	0.8246
Human Observer	0.9473	-
Staal	0.9442	0.9520
Niemeijer	0.9416	0.9294
Zana	0.9377	0.8984
Al-Diri	0.9258	-
Jiang	0.9212	0.9114
Martínez-Pérez	0.9181	-
Chaudhuri	0.8773	0.7878

6.2.5 Chosen Method

After assessing the performance of the different approaches, the SCM method was chosen. This method was superior for almost all metrics, with the exception of the low specificity values on the ROC curve, where Gabor was superior. However, a low specificity is not acceptable, since it may result in the removal of true MA from the candidates.

Another advantage of SCM over Gabor is its computational efficiency, since the application of multiple filters (for each direction) negatively affects Gabor method performance.



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

(e) (f) (g) (h)

Figure 6.13: Comparison between different threshold values ($Threshold = 2$ for top row and $Threshold = 8$ for bottom row) on TP results (images a and e), TN results (images b and f), FP results (images c and g) and FN results (images d and h) for the SCM method. The TP images evidentiate the decrease in detected vessel pixels for the higher threshold.

6.2.6 Effect on removal of false candidates

The effect of removal of candidates located in vessels from the result of the candidate extractor is displayed in table 6.4. The reduction on false positives ranges from 20% to roughly 30%. Simultaneously, the number of removed MA is very low, further proving the effectiveness of this approach.

Table 6.4: Effect of the removal of false candidates detected on vessels, for different numbers of false positives. The candidate extractor used is the and the vessel segmentation method is the SCM. The number of false positives is relative to the mean per image.

False Positive Candidates			**MA Candidates**	
Before removal	**After removal**	**% Removal**	**Removed MA**	**Detected MA**
6.86	5.34	22.16%	1	92
11.52	9.02	21.70%	1	107
25.18	19.02	24.46%	1	129
55.64	40.06	28.00%	2	151
92.26	63.32	31.37%	6	159

6.3 Classification

In order to distinguish between false positive MA candidates and true MA, a classification step is used.

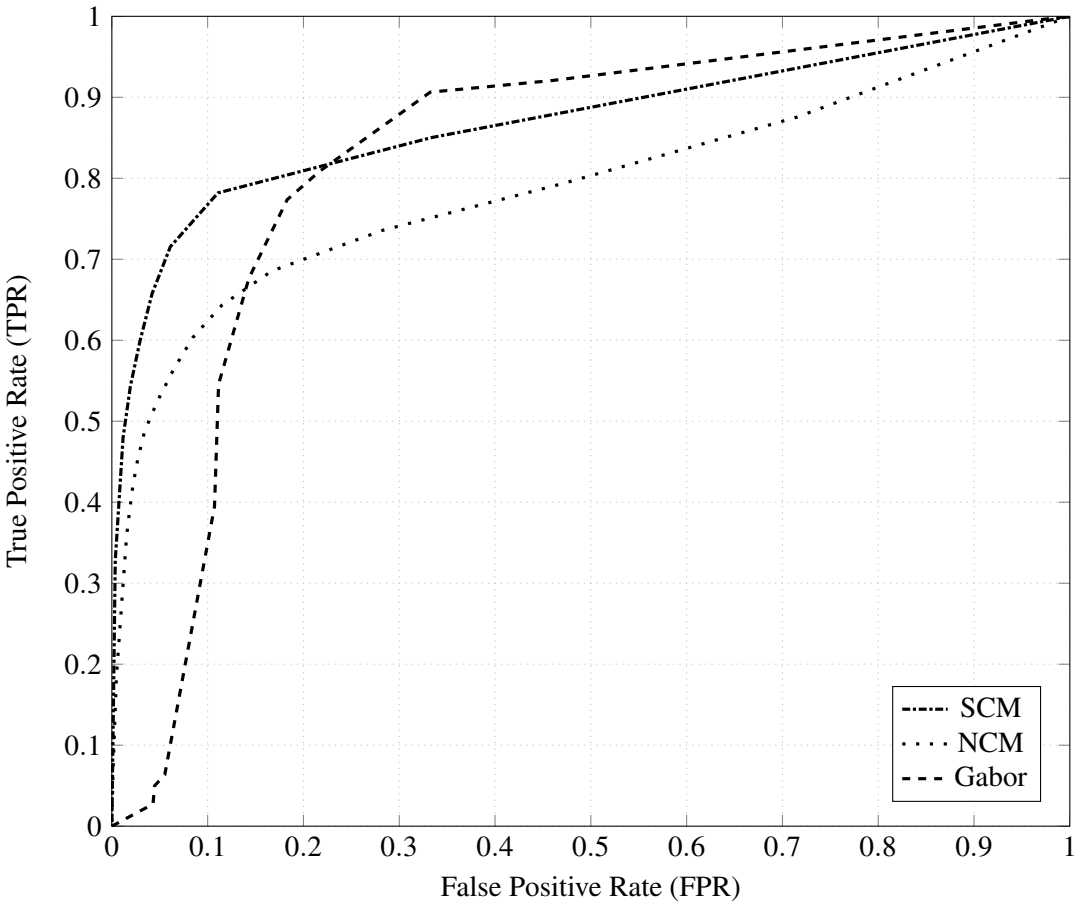


Figure 6.14: ROC curve comparison for the SCM, NCM and Gabor methods.

In order to make the implementation easier, only classifiers supported by OpenCV were considered. Since the OpenCV related processing steps are implemented in the Android application using the Android NDK, employing third party Java libraries to perform the classification would unnecessarily complicate the MA detection pipeline.

6.3.1 Region Growing

After the microaneurysm candidate extraction step, relevant features should be computed in order to perform classification. However, the candidate extraction step only outputs the location of each candidate and no shape information is obtained.

This poses some issues, since most features require shape information, either directly - for features such as the circularity, the area and the aspect ratio of the candidate - or indirectly - for features like the mean green channel intensity or the ratio of the mean intensity of the candidate in relation with the neighbour pixels.

In order to solve this problem, an algorithm rooted on the region growing method was devised. Region growing is a segmentation method which, given a set of initial points (the seeds - in this case, the only seed is the location of the candidate), iteratively adds

its neighbours to the segmented region, provided their intensity difference to that region is below a certain threshold. The region growing procedure is repeated with different, increasing thresholds, stopping when the candidate region has an area larger than what could be expected for a microaneurysm. The flowchart of this algorithm for each candidate is represented in figure 6.15.

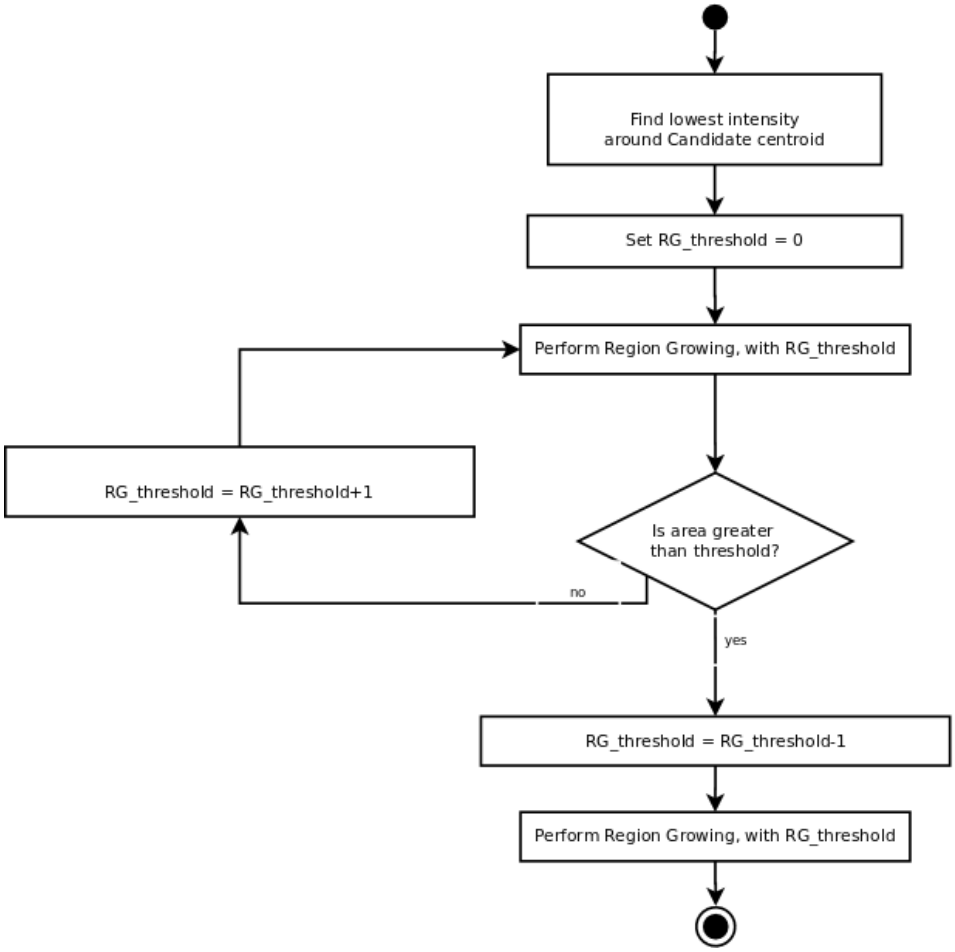


Figure 6.15: Candidate shape estimation algorithm flowchart.

Experimentally, this approach led to good shape estimation and generally the microaneurysm candidates have an area significantly smaller than the maximum admitted. Examples of the area variation with different thresholds are presented in table 6.5.

6.3.2 Features

In order to accurately classify candidates as MAs or not, a set of relevant features should be computed for each of the candidates. Ideally, those features should discriminate true MAs from the remaining candidates, in order to allow the classifier to attribute each candidate to the adequate class (MA or not).

Table 6.5: Example of the microaneurysm candidate area variation with the increase of the region growing threshold.

Candidate	Threshold	Candidate Area
1	1	3
	2	4
	3	4
	4	8
	5	13
	6	111
2	1	2
	2	5
	3	6
	4	10
	5	55
3	1	1
	2	3
	3	8
	4	13
	5	15
	6	200547
4	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	5
	9	6
	10	6
	11	7
	12	9
	13	108

A total of 30 features were used, rooted on previous work by *Spencer 1996* and *Niemeijer 2005* [18, 17]. Other features, directly related with the candidate extraction step, were also used. The features are described below:

Area the area of the candidate region.

Perimeter the perimeter of candidate region.

Aspect ratio aspect ratio, defined as the ratio between the major and the minor axis.

Circularity defined as $4\pi a/p^2$.

Total Intensity sum of the intensities of the candidate region, in the green channel.

Total Intensity Corrected Background total intensity of the candidate region in the background corrected green channel. Background correction corresponds to subtracting the original green channel by its median filtered version.

Mean Intensity mean intensity of the candidate region in the green channel.

Mean Intensity Corrected Background mean intensity of the candidate region in the background corrected green channel.

Normalized Total Intensity sum of the intensities of the candidate in the normalized green channel image. The normalization consists in subtracting the mean of the original image and dividing by the standard deviation.

Normalized Total Intensity Corrected Background sum of the intensities of the candidate in the normalized green channel image with background correction.

Normalized Mean Intensity mean intensity of the candidate region in the normalized green channel.

Normalized Mean Intensity Corrected Background mean intensity of the candidate region in the normalized green channel with corrected background.

Seed Point Intensity intensity of the region growing seed point in the green channel.

Compactness defined as $\sqrt{\sum (d_i - d)^2 / n}$, where d_i is the mean of all the distances between the boundary pixels and the center of the candidate, d is the mean of those distances and n is the number of pixels in the boundary.

Mean Difference Red Channel difference between the mean intensity of the candidate region and the mean intensity of their neighbours, in the red channel.

Mean Difference Green Channel difference between the mean intensity of the candidate region and the mean intensity of their neighbours, in the green channel.

Mean Difference Blue Channel difference between the mean intensity of the candidate region and the mean intensity of their neighbours, in the blue channel.

Mean Difference Hue Channel difference between the mean intensity of the candidate region and the mean intensity of their neighbours, in the hue channel of the HSV color space.

Gaussian Correlation ($\times 5$) correlation value of the candidate, for all values of sigma.

Gaussian filtered mean intensity ($\times 4$) the mean intensity of the candidate in the gaussian filtered green channel, for different gaussian kernels (sigma = 1, 2, 4 and 8).

Standard deviation ($\times 4$) the standard deviation of the candidate in the gaussian filtered green channel, for different gaussian kernels (sigma = 1, 2, 4 and 8).

Since the value range for different features varies widely, classifiers which rely on the distance between samples in the feature space (as is the case of k-NN) could be negatively affected. In order to solve this issue, the feature values are scaled to lie in the range $[-1,1]$:

$$x' = 2 \cdot \frac{x - min(x)}{max(x) - min(x)} - 1$$

6.3.3 Feature Selection

As described in the previous section, the computation of some features requires the application of filtering and other image processing methods that are computationally expensive. Furthermore, the use of a large number of features may lead to overfitting, especially when the size of the training dataset is limited. Another concern is that some features may be irrelevant, not improving the discriminative capabilities of the classifier.

The use of feature selection methods reviewed in section 3.6.3 allow us to address these issues, by obtaining a subset of the original features for classification.

Performing an exhaustive search, by using all the possible combinations of features would guarantee to find the optimal subset of features but is computationally infeasible - for 30 features, the total number of combinations to evaluate would be $N_{Combinations} = 30! = 2.65 \times 10^{32}$. As such, in order to cope with this task, the forward feature selection technique was used.

Two different classifier methods were employed - k-NN and SVM, reviewed in section 3.6 - and feature selection was performed for both classifiers. Since the classes are highly imbalanced (the number of FP candidates is much larger than the number of true MAs), the performance metric chosen was not the accuracy, but the F1 score, which consists of a harmonic mean between the specificity and the sensitivity of the results:

$$F_1 = 2 \cdot \frac{spec * sens}{spec + sens}$$

For the k-NN classifier, the selected features were (by order of addition to the final subset):

- Correlation ($\sigma = 1.7$)
- Mean Difference Hue Channel
- Normalized Mean Intensity
- Gaussian filtered mean intensity ($\sigma = 1$)

- Mean Difference Red Channel
- Compactness
- Correlation ($\sigma = 1.4$)
- Normalized Mean Intensity Corrected Background
- Mean Difference Green Channel

For the SVM classifier, the selected features were (also by order of addition):

- Correlation ($\sigma = 1.4$)
- Normalized Mean Intensity Corrected Background
- Mean Difference Red Channel
- Circularity
- Gaussian filtered mean intensity ($\sigma = 4$)
- Normalized Mean Intensity
- Mean Intensity Corrected Background
- Standard deviation ($\sigma = 1$)
- Total Intensity Corrected Background

6.3.4 Dealing with unbalanced datasets

As concluded in section 6.1.3, the output of the candidate extractor step contains a much larger number of false candidates than true MAs. In fact, for a sensitivity of 41.07% (corresponding to a total of 138 true MAs extracted) the total number of false positives is 1211 - which is approximately a 1:9 TP/FP ratio.

Classification of highly imbalanced datasets poses a challenge, since bias is generally verified for the majority class. In order to attenuate these effects, undersampling is used, which simply consists in discarding a fraction of the majority class samples. Even though this approach is not ideal, since the elimination of the labelled samples may hinder the robustness of the model, it is effective in removing the bias in favour of the majority class.

6.3.5 Results

As previously mentioned, two different classifiers were employed: k-NN and SVM. The final results were obtained using the subset of features described in section 6.3.3.

Since the ground truth for the ROC test dataset was not made available, the results were obtained for the training dataset. However, obtaining results for the same data used for training the classifier would be a methodological error, as it could mask potential abnormalities of the generated model. One of most common abnormalities is overfitting, where the performance of the model on the training data is much higher than registered on unseen observations.

One of the possibilities to avoid this issue is to further split the training set, using one of the sets to train the classifier and the other to test it. However, since the number of available samples in the training set is relatively low (a total of 336 microaneurysms in 50 images), this approach would significantly limit the amount of data available for training and testing.

Another approach is to use the k-fold cross validation technique. It consists in splitting the original dataset (in this case, the ROC training dataset) into k subsets with the same number of samples and TP/FP proportions. Training the classifier is performed using $k-1$ subsets, and results are obtained for the remaining subset. This process is repeated until each of the k subsets is used as a test set. The iterative nature of this process is illustrated in figure 6.16.

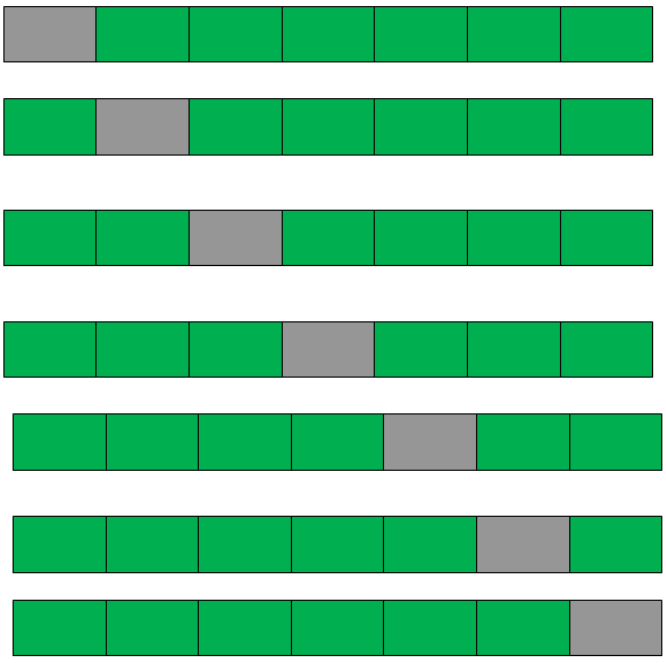


Figure 6.16: Example of the k-fold cross validation technique. In this case, $k = 7$, and all the iterations are represented from the top to the bottom. The test subsets are represented in grey and the training subsets are represented in green.

Results of classification are represented in table 6.6. These results are relative to the entire pipeline, already accounting for the MAs not detected in the candidate extraction step. The number of false positives used to compare the different approaches mimic the methodology used for the comparison in the Retinopathy Online Challenge.

The sensitivity results for the SVM classifier are higher for all number of false positives, even though with only marginal advantage over the *Zhang 2010* method. The sensitivity results of the k-NN classification are worse than the SVM results but on par with the Zhang et al method - for low values of FP per image it performs better, but for higher values it has worse performance. The Math morphology method presents the best results when the average FP per image is $1/8$ but it has worse performance than all the other methods for the remaining values.

Table 6.6: Sensitivity results for the MA detection pipeline, on the ROC training dataset.

FPs / Scan	**1/8**	**1/4**	**1/2**	**1**	**2**	**4**	**8**
GC w/SVM	6.2%	10.8%	14.83%	19.6%	27.8%	32.4%	35.8%
GC w/k-NN	6.8%	10.5%	12.6%	15.7%	23.8%	28.7%	34.7%
Zhang et al	6.0%	9.6%	13.8%	18.9%	26.7%	31.1%	34.5%
Math Morp	8.1%	8.7%	10.7%	10.7%	14.1%	16.6%	19.6%

Chapter 7

Android Application

An Android application was developed, in order to allow the acquisition of retinal images, management of the patient data and detection of microaneurysms in the acquired images.

The application uses the design guidelines defined by the SmartCompanion project [49], which seeks to attend senior citizens needs with an easy to use smartphone experience. The design language includes high contrast items, well defined navigation patterns (back button always available on the action bar, action buttons on the bottom of the screen) and the use of large size fonts.

7.1 Database

In order to persist and manage the data introduced by the user, a local SQLite database is used. The entity relationship (ER) diagram of the database is represented in figure 7.1.

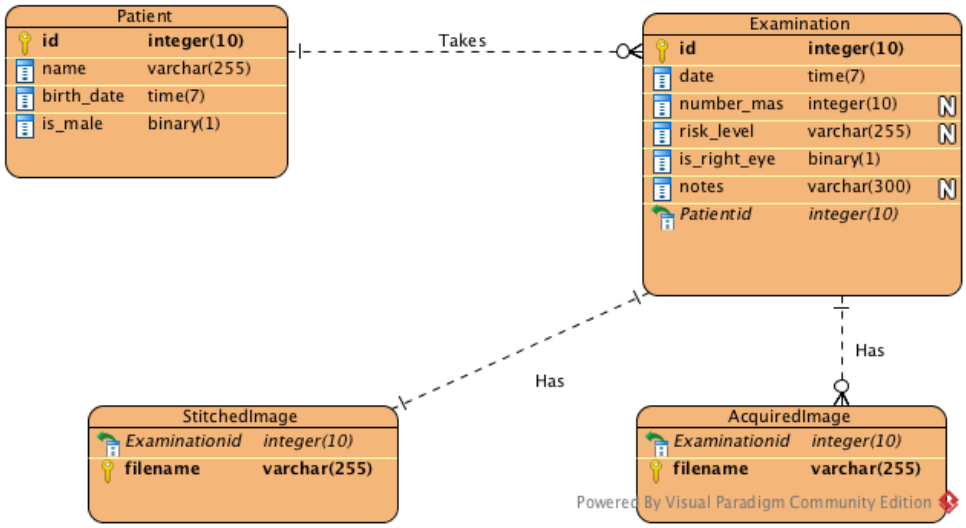


Figure 7.1: Entity relationship diagram of the SQLite database.

Four different entities are present: *Patient*, *Examination*, *AcquiredImage* and *StitchedImage*. The *Patient* entity has a one to many relationship with the *Examination* (a patient may take multiple examinations). On the other hand, the *Examination* entity has a one to many relationship with the *AcquiredImage* entity and a one to one relationship with the *StitchedImage* entity (for each examination, only a single stitched image is considered for analysis).

In order to abstract lower level details of the database implementation, Data Acess Object (DAO) classes were created with static methods that perform the needed operations - create a patient in the database, retrieve a specific exam given its *id*, retrieve all examinations, obtain the file names of every acquired image given the *id* of the examination, etc.

7.2 Navigation

The navigation map of the application is represented in figure 7.2.

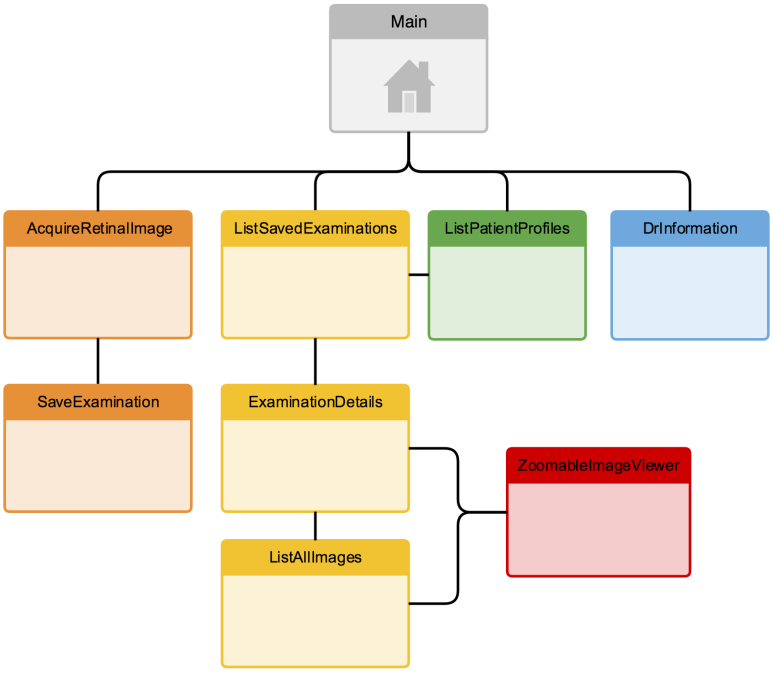


Figure 7.2: Navigation between different activities in the developed Android application.

The *Main* activity only displays four possible options, through buttons, to the user: to acquire new retinal images, to view saved examinations, to view a list of the user profiles or to consult information about DR. The layout of this activity is represented in figure 7.3

If the user decides to take a retinal image, a new activity (*AcquireRetinalImage*) is started, from which the patient is shown a preview of the camera and, at any time, may

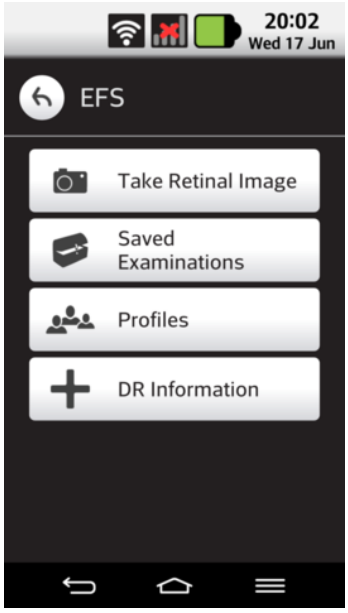


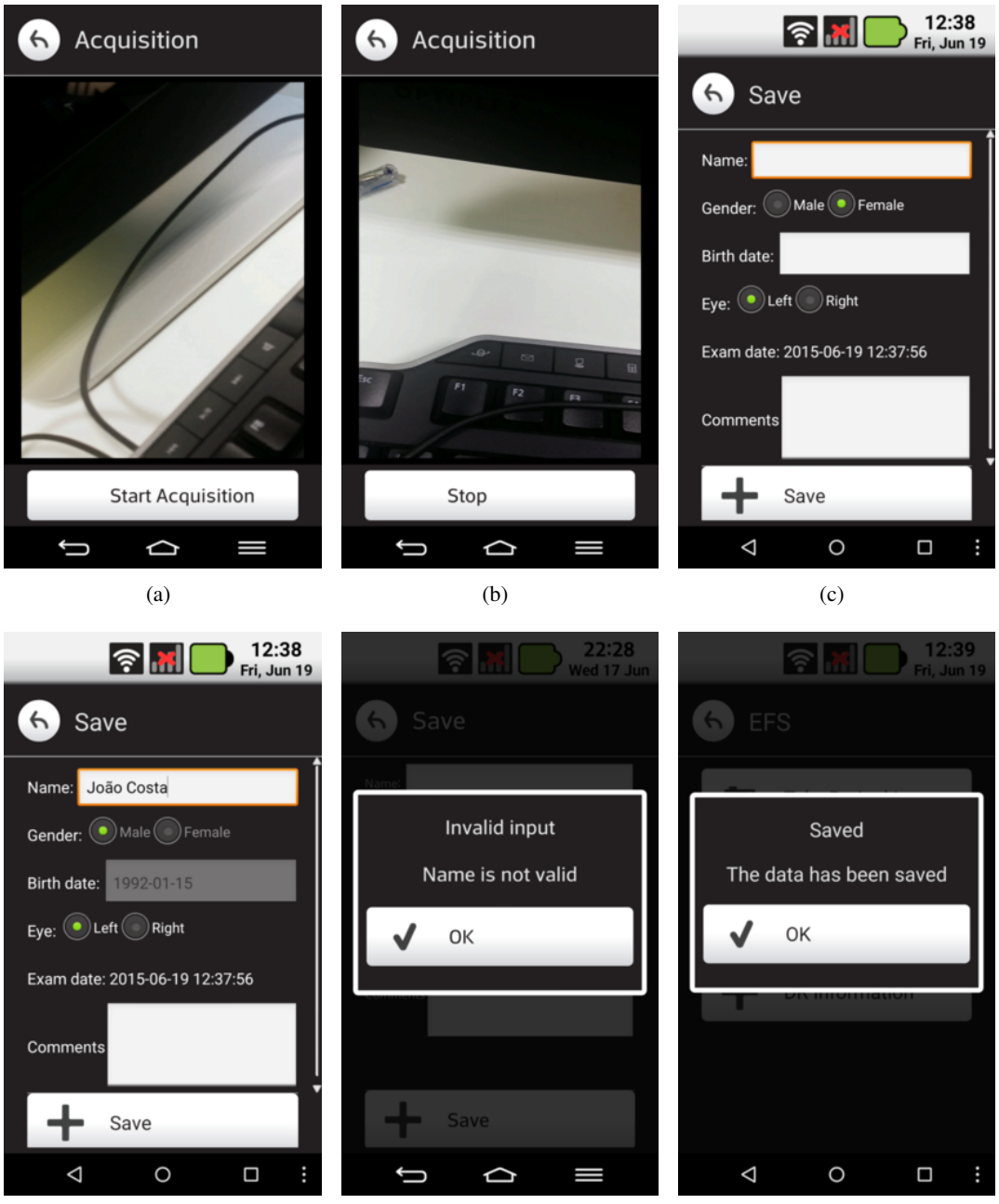
Figure 7.3: Layout of the *Main* activity.

start the examination. After the examination is finished, another activity (*SaveExamination*) is started, allowing the user to save relevant details. Before the examination is saved, the user input is validated and in case it is not acceptable, a warning dialog is shown and the user is asked to try again. Please refer to figure 7.4 for the layouts of the use cases described.

In the *ListSavedExaminations* activity, all the examinations registered in the database are displayed to the user, including appropriate information. If the user clicks one of the particular examinations, a new activity (*ExaminationDetails*) is started, presenting the examination details. In the top, the stitched image is displayed, followed by the complete information of the examination, including the risk level and number of MAs detected. From this activity, the user may select a button and either delete the examination or view all the acquired images. For the latter, a new activity (*ListAllImages*) is once again started, displaying a grid of all the images captured. Either on the *ExaminationDetails* or in the *ListAllImages* activities the user may click one of the images and view it in complete detail with possibility of performing zoom (in the *ZoomableImageViewer* activity). The layouts of the different activities and use cases are illustrated in figure 7.5.

From the *Main* activity, if the user selects the "Profiles" button, the *ListPatientProfiles* activity is started, displaying all the available patient profiles, along with relevant information. If the user selects one of the list items, the examinations performed for the respective patient are shown in a new activity. The layouts for these use cases are shown in figure 7.6.

From the *Main* activity, if the user selects the "DR Information" button, the *DrInformation* activity is started, showing the user some information about diabetic retinopathy



(a) (b) (c)

(d) (e) (f)

Figure 7.4: Retinal image acquisition and exam saving form layouts and use cases.

(see figure 7.7).

7.3 Acquisition

The acquisition process takes place in the *AcquireRetinalImage* activity. The acquisition is started when the user presses the "Start Acquisition" button, and stops when the user

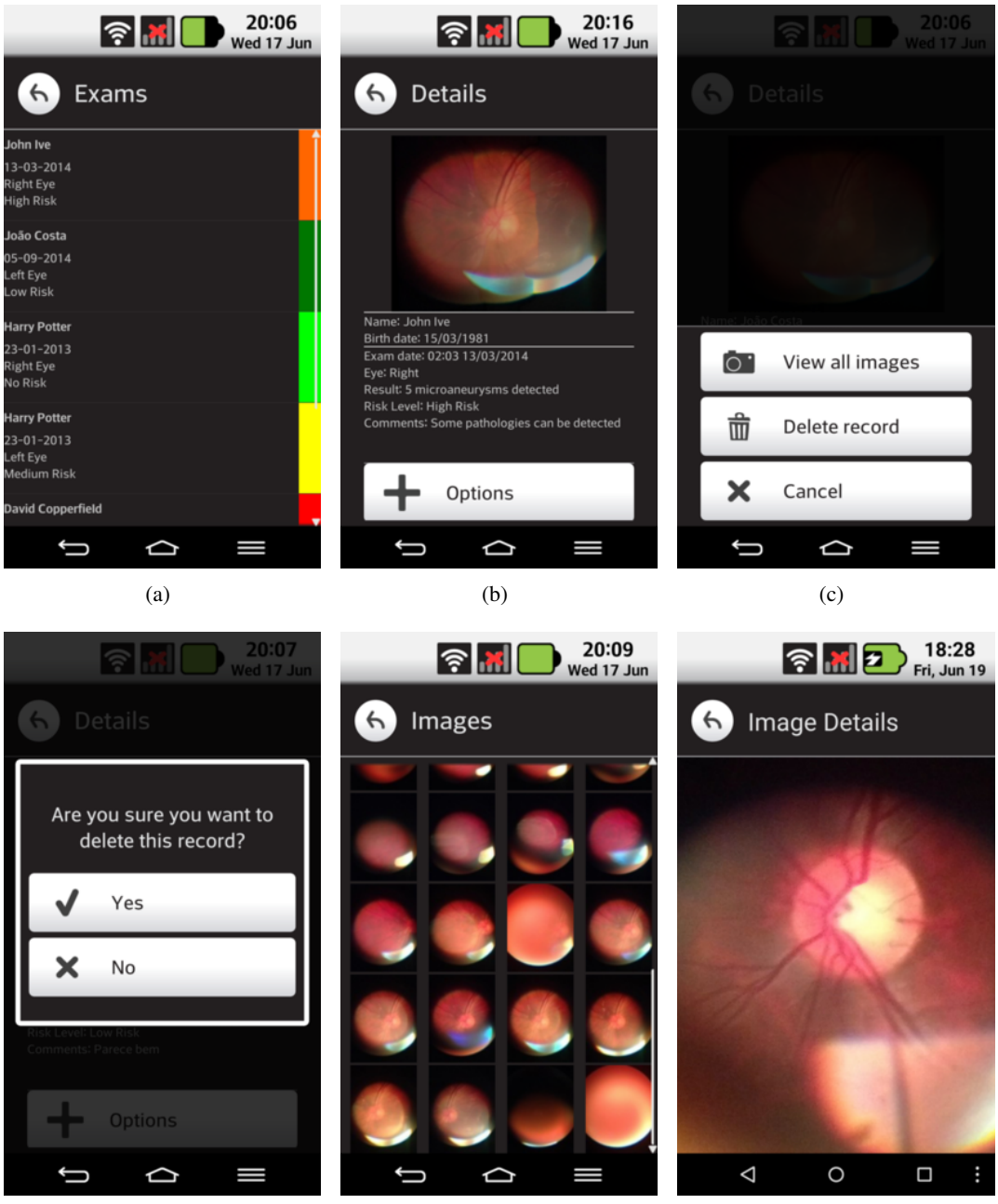
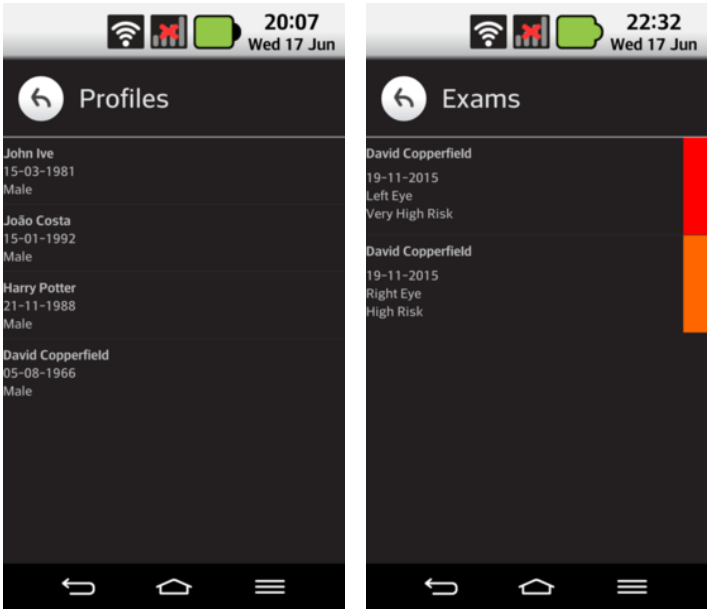


Figure 7.5: *ListSavedExaminations* and related activities layouts and use cases.

presses the "Stop" button.

An image is acquired each 0.5 seconds and saved to a file using the Android *AsyncTask* class. The list of the file names of the acquired images is persisted in the database, and this list is used to retrieve the images to perform the image stitching procedure.

The acquisition methodology is described in the the acquisition protocol (in annex B). In figure 7.8, an example of the acquisition is shown.



(a) (b)

Figure 7.6: *ListPatientProfiles* activity and *ListExaminations* activity layouts.

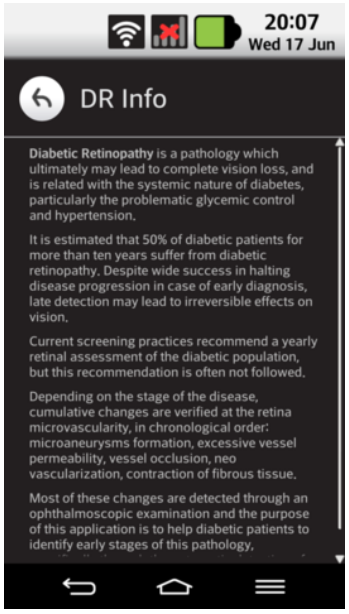


Figure 7.7: *DrInformation* activity layout.

7.4 Microaneurysm Detection Algorithm Integration

The MA detection pipeline is integrated using the Android Native Development Kit. This allows the original developed C++ code to be employed without any major modifications.

The processing operations occur after the user presses the "Save" button on the *SaveExamination* activity. These operations are performed in a different thread, using the

Figure 7.8: Patient and user during acquisition.

Android *Service* component. The MA detection is performed on the stitched image and, as such, it is only started after the computation of this image.

Chapter 8

Conclusion

Diabetic retinopathy is the major cause of blindness for people between 20 and 74 years old and even though yearly screening for the diabetic population is recommended, often this is not followed. The asymptomatic nature of the initial stages of diabetic retinopathy contribute to this problem, and often it is only detected were changes are already irreversible.

In this work, a mobile solution to perform automatic preliminary screening of diabetic retinopathy was developed. In order to make this possible, a set of objectives were set:

- Definition of an image acquisition protocol for smartphone fundography.
- Computer aided alignment of the eye during smartphone image acquisition.
- Increase available retinal FOV through image stitching.
- Image recognition and automatic annotation of microaneurysms on eye fundus.

An image acquisition protocol was in fact devised (please refer to appendix B) to standardize the acquisition process. This protocol revealed itself to be useful to improve the quality of the acquired images and the ease of use of the system, even though new users still faced some initial difficulties with the acquisition.

An algorithm to perform alignment of the eye was also developed, through the detection of the optic disc. This algorithm was validated on public datasets and successfully tested in mobile acquired images.

Through the use of image stitching, the FOV of the retinal images was increased, allowing for a more robust assessment of the patient condition. This step is particularly important in order to improve the quality of the examination, since performing microaneurysm detection on a small retinal area would limit the reliability of the results.

Finally, a complete microaneurysm detection method was developed, with several processing stages. These include the extraction of candidates through a template matching approach, the segmentation of the retinal vessels and removal of false candidates present

there and finally a classification of the candidates as microaneurysms or not microaneurysms, using a set of optimal features. The overall performance of this pipeline is on par with other state of the art approaches.

The major breakthrough of this project is the development of an Android application capable of dealing with the acquisition and enhancement (through stitching) of retinal images, patient data management and the detection of microaneurysms in those images. This application integrates the methods defined in the objectives for this thesis.

8.1 Future Work

During the development of this thesis, several aspects were noted which could improve the general objectives of the project, but whose development could not be met within the given timeframe. The most relevant of these aspects are described in the following sections.

8.1.1 Validation on different datasets and smartphone acquired images

While the developed microaneurysm detection algorithm was successfully tested on the ROC dataset, performing further evaluation on different data could help in its validation. One example of another possible dataset is DIARETDB1 [48]: similar to ROC, this dataset provides lesion locations as its ground truth, but includes other lesions besides microaneurysms.

Validation of the MA detection algorithm in smartphone acquired images is also extremely important, but could not be performed in the timeframe of this thesis. Besides the acquisition of the images (preferably, in a DR screening program), performing this type of validation would require annotation of the images by experts.

8.1.2 Detection of exudates

As detailed in section 2.2, microaneurysms are the first detectable signs of DR, but they are not the only ones. Exudates are often formed as a result of slightly more advanced stages of the pathology, and their detection is used for the evaluation of the patient condition.

As a future step, the detection of exudates should also be performed, in order to provide a more robust system.

8.1.3 Multi image denoising

Due to the fact that the acquisition of retinal images takes place in a low light setting, substantial noise is often present, as a result of the high ISOs employed. The high alignment

accuracy obtained using direct alignment can be employed to improve the image quality by reducing noise. The acquisition use case is particularly suited for this task, since it involves capturing a large number of images in a short timespan and with a substantial overlap.

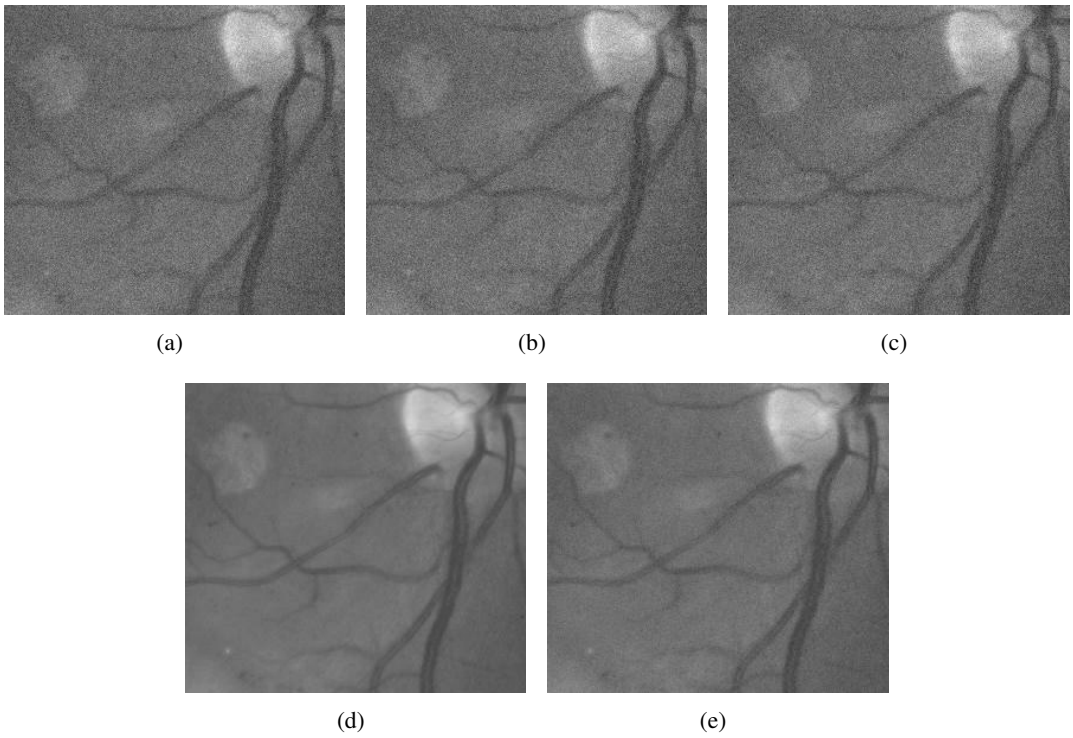


Figure 8.1: Image averaging for noise reduction. (a,b,c) Input noisy images; (d) Original image; (e) Result of averaging.

Simply averaging the pixel intensities of aligned overlapping images is often used to reduce noise (please refer to figure 8.1). More advanced techniques for multi image denoising generally produce better results if the unprocessed RAW images are used, as these images preserve the full bit depth of the sensors [50, 51]. From Android 5.0, it is possible to obtain these RAW images in compatible smartphones - at the time of writing, only LG Nexus 5, LG Nexus 6 and HTC One M9.

Multi image denoising is particular advantageous for the case of detecting microaneurysms, since using single image denoising approaches such as linear low pass filters or median filters may mask existent MAs.

8.1.4 Near Infrared ophthalmoscopy

One of the great challenges of mobile fundoscopy is the requirement of an intense light source. Not only does this cause discomfort for the patient, but also causes a pupillary response (in non mydriatic settings), which diminishes the FOV available and increases

the difficulty of the acquisition. This issue is somehow avoided in desktop cameras by only using the light source when the image is actually taken (using a flash) - this is possible because the alignment between the patient pupil and the optic system is better controlled in these conditions.

A possible solution to improve the acquisition quality and ease of use in a mobile setting is to use a near infrared light source. Since the eye is not sensible to the wavelengths used, the patient will not feel discomfort and pupillary response will be avoided.

Variations of infrared ophthalmoscopy are already used in clinical settings, for desktop fundoscopy cameras [52]. Using this technique in smartphone based acquisition would require, besides the adaptation of the light source, some modifications in the smarthphone camera module, in order to improve the camera sensitivity for lower wavelengths.

8.1.5 Cloud based processing

Offloading the computationally intense image processing stages to a remote server could allow the use of more advanced algorithms for MA detection.

One example of an algorithm that could be used on this context is spectral clustering. Even though it was initially proposed, it was verified that it was inadequately heavy to be employed on a mobile context.

One problem with using cloud based processing is that this approach significantly increases the complexity of the project and would require an internet connection on the part of the user.

Appendix A

Difference between different channels

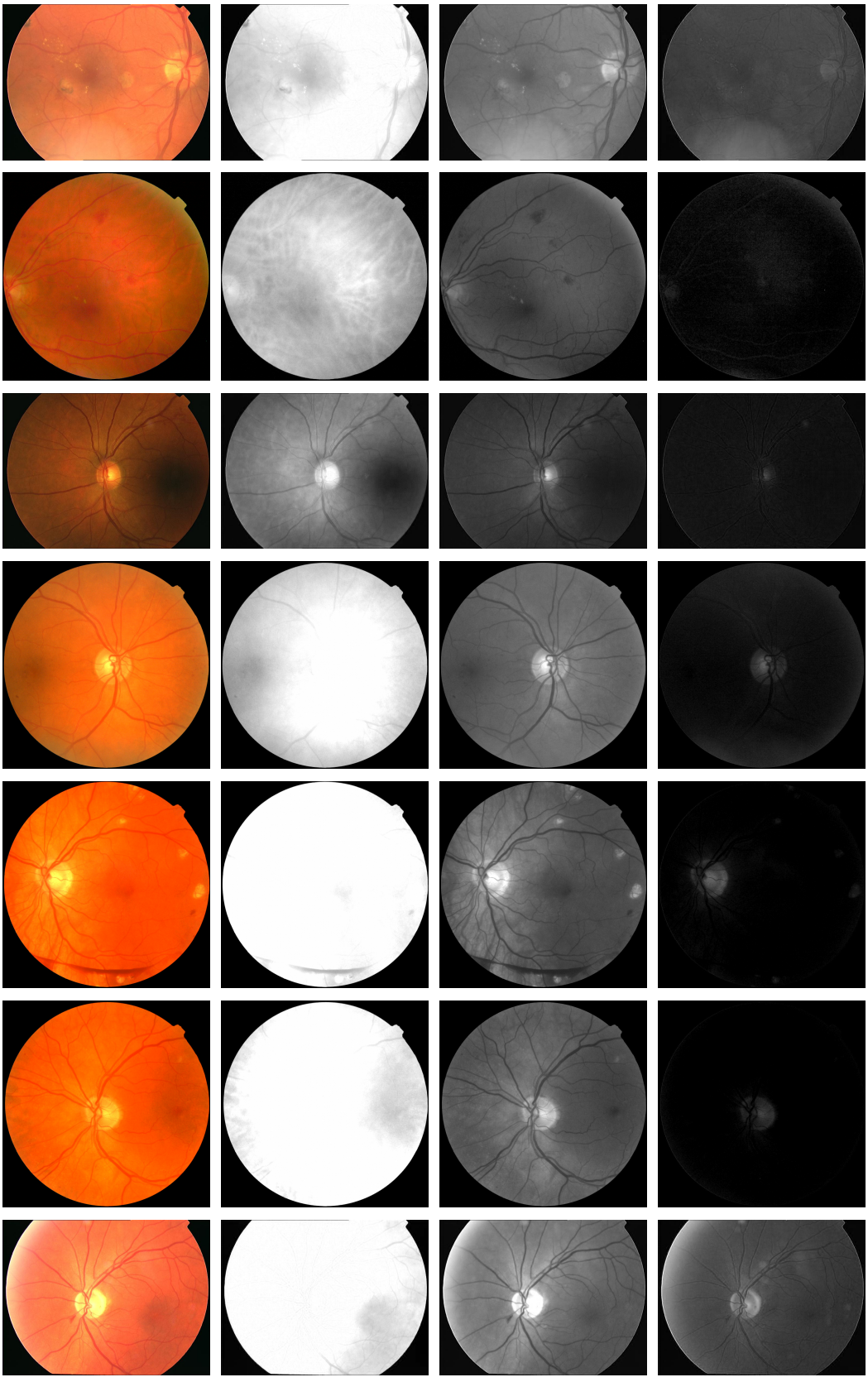


Figure A.1: Differences between channels in retinal images, for the RGB color space. In the first column the color images are represented, in the second column is the red channel, in the third column is the green channel and in the fourth column is the blue channel.

Appendix B

Acquisition Protocol

EyeFundusScope Image Acquisition Protocol

The goal of this acquisition protocol is to standardize the acquired images, and ensure its quality.

Before starting the examination, some conditions should be considered. First, for the sake of image quality, ambient light should be reduced as much as possible, and so the room where the acquisition is performed should be dark.

Pupil dilation is not necessary for the acquisition with the *Welch Allyn* PanOptic ophthalmoscope, but its use could lead to an increase of the Field of View, as well as to facilitate some steps of this protocol.

Patient Instructions

1. Sit across the patient, in the diagonal with the eye that will be examined.
2. Instruct the patient that he may blink as needed, but otherwise should try to keep his eyes as open as possible.
3. Ask the patient to focus, for the entire duration of the exam, in a dark spot (for example, in the wall). Ask to not focus on the ophthalmoscope light.
4. Explain that the intensity of the light may be slightly uncomfortable, but the expected duration of the examination is under 50 seconds.

Ophthalmoscope Adjustment

1. Adjust the light intensity using the rheostat to half the maximum intensity.
2. Change the ophthalmoscope aperture to small (home position, marked with a green mark).
3. Adjust the focus wheel around zero - if the patient has refractive error, it will be adjusted by the automatic focusing performed by the smartphone camera.

Examination

1. Begin the acquisition.

2. Start with the ophthalmoscope distanced from the patient eye around 10 cm, in a 20 angle, temporally, on the same horizontal plane.

3. The red reflex should be immediately noticeable. In case it is not change the angle slightly until it is visible.

4. Follow the red reflex, until the image is focused and the retina is clearly visible. In order to capture different retinal regions, change the angle progressively, until approximately a 120 angle.

5. Stop the acquisition.

6. Repeat for the other eye.




References

[1] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.

[2] Eva M Kohner, M Sleightholm, Kroc Collaborative Study Group, et al. Does microaneurysm count reflect severity of early diabetic retinopathy? *Ophthalmology*, 93(5):586–589, 1986.

[3] Ronald Klein, Stacy M Meuer, Scot E Moss, and Barbara EK Klein. Retinal microaneurysm counts and 10-year progression of diabetic retinopathy. *Archives of Ophthalmology*, 113(11):1386–1391, 1995.

[4] Jie Ding and TienYin Wong. Current epidemiology of diabetic retinopathy and diabetic macular edema. *Current Diabetes Reports*, 12(4):346–354, 2012.

[5] Cheung N, Mitchell P, and Wong TY. Diabetic retinopathy. *Lancet*, 10(376):124–36, 2010.

[6] F. Oloumi, R.M. Rangayyan, and A.L. Ells. *Digital Image Processing for Ophthalmology: Detection and Modeling of Retinal Vascular Architecture*. G - Reference,Information and Interdisciplinary Subjects Series. Morgan & Claypool, 2014.

[7] D. LeRoith, S.I. Taylor, and J.M. Olefsky. *Diabetes Mellitus: A Fundamental and Clinical Text*. M - Medicine Series. Lippincott Williams & Wilkins, 2004.

[8] Serge Resnikoff, Donatella Pascolini, Daniel Etya'ale, Ivo Kocur, Ramachandra Pararajasegaram, Gopal P. Pokharel, and Silvio P. Mariotti. Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization*, 82:844 – 851, 11 2004.

[9] Lawrence A. Yannuzzi, Kathleen T. Rohrer, Lori J. Tindel, Russell S. Sobel, Marcelle A. Costanza, William Shields, and Edith Zang. Fluorescein angiography complication survey. *Ophthalmology*, 93(5):611 – 617, 1986.

[10] MD Veronica Kon Graversen, MD; Pooja Jani. Clinically useful and cost-effective mobile retinal screening. *New Retina MD*, 2014.

[11] William Tasman and Edward A. Jaeger. *Duane's Ophthalmology*. 2009.

[12] István Lázár and András Hajdu. Microaneurysm detection in retinal images using a rotating cross-section based model. In *ISBI*, pages 1405–1409. IEEE, 2011.

[13] A M Mendonca, A J Campilho, and J M Nunes. Automatic segmentation of microaneurysms in retinal angiograms of diabetic patient. In *in Proc. Int. Confe. Image Anal. Process*, pages 728–733, 1999.

[14] CE Baudoin, BJ Lay, and JC Klein. Automatic detection of microaneurysms in diabetic fluorescein angiography. *Revue d'epidemiologie et de sante publique*, 32(3-4):254—261, 1984.

[15] Ana Maria Mendonca and Aurelio Campilho. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *Medical Imaging, IEEE Transactions on*, 25(9):1200–1213, 2006.

[16] A. Agrawal, C. Bhatnagar, and A.S. Jalal. A survey on automated microaneurysm detection in diabetic retinopathy retinal images. In *Information Systems and Computer Networks (ISCON), 2013 International Conference on*, pages 24–29, March 2013.

[17] Meindert Niemeijer, Bram van Ginneken, Joes Staal, Maria S. A. Suttorp-Schulten, and Michael D. Abràmoff. Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans. Med. Imaging*, 24(5):584–592, 2005.

[18] Timothy Spencer, John A. Olson, Kenneth C. McHardy, Peter F. Sharp, and John V. Forrester. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput. Biomed. Res.*, 29(4):284–302, August 1996.

[19] M. Garcia, C.I. Sanchez, Maria I. Lopez, Ana Diez, and R. Hornero. Automatic detection of red lesions in retinal images using a multilayer perceptron neural network. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 5425–5428, Aug 2008.

[20] María García, María I López, Daniel Álvarez, and Roberto Hornero. Assessment of four neural network based classifiers to automatically detect red lesions in retinal images. *Medical engineering & physics*, 32(10):1085–1093, 2010.

[21] S. Balasubramanian, S. Pradhan, and V. Chandrasekaran. Red lesions detection in digital fundus images. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 2932–2935, Oct 2008.

[22] Y. Hatanaka, T. Inoue, S. Okumura, C. Muramatsu, and H. Fujita. Automated microaneurysm detection method based on double-ring filter and feature analysis in retinal fundus images. In *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, pages 1–4, June 2012.

[23] iain livingstone, Andrew Bastawrous, Mario E Giardini, Stewart Jordan, and Peek Collaboration. Peek: Portable eye examination kit. the smartphone ophthalmoscope. *Invest. Ophthalmol. Vis. Sci.*, 55(5):1612, 2014.

[24] Meindert Niemeijer, Bram Van Ginneken, Michael J Cree, Atsushi Mizutani, Gwénolé Quellec, Clara Sánchez, Boming Zhang, Roberto Hornero, Mathieu

Lamard, Chisako Muramatsu, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *Medical Imaging, IEEE Transactions on*, 29(1):185–195, 2010.

[25] Richard Szeliski. Image alignment and stitching: A tutorial. Technical Report MSR-TR-2004-92, Microsoft Research, October 2004.

[26] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008. Similarity Matching in Computer Vision and Multimedia.

[27] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of the 2011 International Conference on Computer Vision*, ICCV '11, pages 2564–2571, Washington, DC, USA, 2011. IEEE Computer Society.

[28] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.

[29] Richard Hartley. Optimised kd-trees for fast image descriptor matching. In *In CVPR, IEEE Computer Society*, 2008.

[30] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[31] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. pages 674–679, 1981.

[32] Qi Tian and Michael N Huhns. Algorithms for subpixel registration. *Comput. Vision Graph. Image Process.*, 35(2):220–233, August 1986.

[33] Alfonso Beato. Opencv 3.0.0 pixel-intensity based registration module, 2014.

[34] Matthew Brown and DavidG. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

[35] H. Yu, E.S. Barriga, C. Agurto, S. Echegaray, M.S. Pattichis, W. Bauman, and P. Soliz. Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. *Information Technology in Biomedicine, IEEE Transactions on*, 16(4):644–657, July 2012.

[36] Marc Lalonde, Mario Beaulieu, and Langis Gagnon. Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. *Medical Imaging, IEEE Transactions on*, 20(11):1193–1200, 2001.

[37] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[38] T. Chanwimaluang, Guoliang Fan, and S. R. Fransen. Hybrid retinal image registration. *Trans. Info. Tech. Biomed.*, 10(1):129–142, January 2006.

[39] Luo Juan and Oubong Gwon. A Comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing (IJIP)*, 3(4):143–152, 2009.

[40] Christoffer Valgren and Achim J. Lilienthal. Sift, surf and seasons: Long-term outdoor localization using local features. In *EMCR*, 2007.

[41] S. Abdelazeem. Micro-aneurysm detection using vessels removal and circular hough transform. In *Radio Science Conference, 2002. (NRSC 2002). Proceedings of the Nineteenth National*, pages 421–426, 2002.

[42] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.

[43] Bob Zhang, Xiangqian Wu, Jane You, Qin Li, and Fakhri Karray. Detection of microaneurysms using multi-scale correlation coefficients. *Pattern Recogn.*, 43(6):2237–2248, June 2010.

[44] Balint Antal and Andras Hajdu. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *Biomedical Engineering, IEEE Transactions on*, 59(6):1720–1726, 2012.

[45] Kedir M Adal, Désiré Sidibé, Sharib Ali, Edward Chaum, Thomas P Karnowski, and Fabrice Mériaudeau. Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning. *Computer methods and programs in biomedicine*, 114(1):1–10, 2014.

[46] Thomas Walter, Pascale Massin, Ali Erginay, Richard Ordonez, Clotilde Jeulin, and Jean-Claude Klein. Automatic detection of microaneurysms in color fundus images. *Medical image analysis*, 11(6):555–566, 2007.

[47] Atsushi Mizutani, Chisako Muramatsu, Yuji Hatanaka, Shinsuke Suemori, Takeshi Hara, and Hiroshi Fujita. Automated microaneurysm detection method based on double ring filter in retinal fundus images. In *SPIE medical imaging*, pages 72601N–72601N. International Society for Optics and Photonics, 2009.

[48] Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen, Lasse Lensu, Iiris Sorri, Asta Raninen, Raija Voutilainen, Hannu Uusitalo, Heikki Kälviäinen, and Juhani Pietilä. The diaretdb1 diabetic retinopathy database and evaluation protocol. In *BMVC*, pages 1–10, 2007.

[49] Smartcompanion.projects.fraunhofer.pt. Smart companion, 2015.

[50] Li Zhang, Alok Deshpande, and Xin Chen. Denoising vs. deblurring: Hdr imaging techniques using moving cameras. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 522–529. IEEE, 2010.

[51] Antoni Buades, Yifei Lou, Jean-Michel Morel, and Zhongwei Tang. Multi image noise estimation and denoising. *preprint, Aug*, 2010.

[52] Ann E Elsner, Stephen A Burns, John J Weiter, and Francois C Delori. Infrared imaging of sub-retinal structures in the human ocular fundus. *Vision research*, 36(1):191–205, 1996.