UNIVERSITE DE BOURGOGNE
U.F.R. SCIENCES ET TECHNIQUES

THESE

Présentée et soutenue publiquement pour l'obtention du grade de

Docteur de l'université

Spécialité : Instrumentation et Informatique de l'Image

par

Luca Giancardo

le 27 Septembre 2011

Automated Fundus Images Analysis Techniques to Screen Retinal Diseases in Diabetic Patients

---

Directeur de thèse: Prof. Fabrice Mériaudeau
Professeur - Université de Bourgogne, France
Co-directeur de thèse: Thomas P. Karnowski
Research Scientist - Oak Ridge National Laboratory, USA

JURY

Rapporteurs :

Prof. Christian Roux	Professeur - Telecome Bretagne, France
Prof. Philippe Bolon	Professeur - Université de Savoie, France

Examinateurs:

Prof. Alfredo Ruggeri	Associate Professor - Università di Padova, Italie
Dr. Meindert Niemeijer	Research Scientist - University of Iowa, USA
Dr. Pierrick Bourgeat	Research Scientist - CSIRO, Australie
Prof. David Fofi	Professeur - Université de Bourgogne, France

Résumé :

Cette thèse a pour objet l'étude de nouvelles méthodes de traitement d'image appliquées à l'analyse d'images numériques du fond d'œil de patients diabétiques. En particulier, nous nous sommes concentrés sur le développement algorithmique supportant un système de dépistage automatique de la rétinopathie diabétique. Les techniques présentées dans ce document peuvent être classées en trois catégories: (1) l'évaluation et l'amélioration de la qualité d'image, (2) la segmentation des lésions, et (3) le diagnostic. Pour la première catégorie, nous présentons un algorithme rapide permettant l'estimation numérique de la qualité d'une seule image à partir de caractéristiques extraites de la vascularisation et de la couleur du fond d'œil. De plus, nous démontrons qu'il est possible d'augmenter la qualité des images et de supprimer les artefacts de réflexion en fusionnant les informations extraites de plusieurs images d'un même fond d'œil (images capturées en changeant le point d'attention regardé par le patient). Pour la deuxième catégorie, deux familles de lésion sont ciblées: les exsudats et les microanévrysmes. Deux nouveaux algorithmes pour l'analyse des images du fond d'œil sont proposés et comparés avec les techniques existantes afin de démontrer leur efficacité. Dans le cas des microanévrismes, une nouvelle méthode basée sur la transformée de Radon a été développée. Dans la dernière catégorie, nous présentons un algorithme permettant de diagnostiquer la rétinopathie diabétique et les œdèmes maculaires en analysant les lésions détectées par segmentation d'image; à partir d'une seule image, notre algorithme permet de diagnostiquer une rétinopathie diabétique et/ou un œdème maculaire en $\sim$ 22 secondes sur une machine à 1,6 GHz avec 4 Go de RAM; de plus, nous montrons les premiers résultats de notre algorithme de détection d'œdème maculaire basé sur des images du fond d'œil multiples, qui peut éventuellement permettre d'identifier le gonflement de la macula même si aucune lésion n'est visible.

Mots clés : analyse du fond d'œil, rétinopathie diabétique, œdème maculaire

---

Abstract :

In this Ph.D. thesis, we study new methods to analyse digital fundus images of diabetic patients. In particular, we concentrate on the development of the algorithmic components of an automatic screening system for diabetic retinopathy. The techniques developed can be categorized in: quality assessment and improvement, lesion segmentation and diagnosis. For the first category, we present a fast algorithm to numerically estimate the quality of a single image by employing vasculature and colour-based features; additionally, we show how it is possible to increase the image quality and remove reflection artefacts by merging information gathered in multiple fundus images (which are captured by changing the stare point of the patient). For the second category, two families of lesion are targeted: exudate and microaneurysms; two new algorithms which work on single fundus images are proposed and compared with existing techniques in order to prove their efficacy; in the microaneurysms case, a new Radon transform-based operator was developed. In the last diagnosis category, we have developed an algorithm that diagnoses diabetic retinopathy and diabetic macular edema based on the lesions segmented; starting from a single unseen image, our algorithm can generate a diabetic retinopathy and macular edema diagnosis in $\sim$22 seconds on a

1.6 GHz machine with 4 GB of RAM; additionally, we show the first results of a macular edema detection algorithm based on multiple fundus images, which can potentially identify the swelling of the macula even when no lesions are visible.

Keywords : fundus image analysis, diabetic retinopathy, macular edema

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

First, I would like to thank my two supervisors, Prof. Fabrice Meriaudeau and Thomas P. Karnowski. They allowed me to begin this PhD and accompanied me towards its completion by fighting a superhuman quantity of bureaucracy, for which I am really grateful. They were always supportive of all my ideas, even if that meant more work for them, something that is really rare to find.

I would like to thank Prof.Christian Roux and Prof. Philippe Bolon for having accepted to review this thesis and for their useful inputs. Also, I would like to thank all the members of my PhD jury who managed to take time off their busy schedules.

Je voudrais remercier le Conseil régional de Bourgogne pour le soutien financier de ce doctorat.

I would like to thank Kenneth J. Tobin, Dr. Edward Chaum, Yaqin Li and Hector Santos Villalobos for their precious collaboration and input.

I would like to thank all the people of the ISML (ISMV) group at the Oak Ridge National Laboratory. They were able to cheer me up in those "moments of desperation" that ever so often happen during a PhD. A special mention goes to Chris, Vincent and Janet that were there from my day one in the "New World".

A big thanks to all my Knovillian music-playing house mates, neighbours and friends, that shared with me some unforgettable jam sessions and gigs. Thank you Eugene, Beavers, Kukuli, Sheldon, JP, Steve, Gaute and Andy, it is also because of you that I can consider Knoxville as another home.

Un ringraziamento va ai Prof. Alfredo Ruggeri ed Enrico Grisan, che mi hanno accolto a "scatola chiusa" nel BioImLab permettendomi di avere una splendida esperienza all'Università di Padova. Un ringraziamento speciale ad Enea, Diego, Lara, Elisa, Parnian ed Anna che mi hanno sopportato in ufficio durante i caldi pomeriggi patavini. Inoltre, non posso dimenticare tutti gli altri dottorandi ed assegnisti del dipartimento di bioingegneria che mi hanno fatto sentire come se fossi stato a Padova da sempre. Ovviamente, un grazie all'intera "Cochabamba Orchestra" passata, presente e futura con la quale abbiamo valorosamente conquistato Aurocastro a suon di concerti.

Un enorme grazie va a tutti componenti di "Casa Durazzo" che mi hanno motivato durante i momenti più difficili. Una menzione speciale va senza dubbio al Dr. Ganesh.

Looking back on these three years, an incredible number of people helped and supported me during my "PhD journey". I would really like to personally thank them one by one, in order to make them understand that even if I had to move away, I will always consider them as great friends.

Infine, un ringraziamento a tutta la mia famiglia ed in particolare a Zia Grazia e Zia Iole.

ENVIRONNEMENTS, SANTÉ, STIC
ECOLE DOCTORALE
UNIVERSITÉ DE BOURGOGNE

Chapter 1

Introduction

Diabetes is a widespread disease in the industrialized and developing world. Centers for Disease Control and Prevention (2011) estimate that by 2030,439 million people will have some form of diabetes world wide. This disease has many adverse effects on the eyes, nervous system, heart, kidneys and other organs, but the most likely to be affected first is the retina, hence the patient' sight. In the United States, diabetes is the leading cause of new cases of blindness among adults aged between 20 and 74 years old. Between 2005 and 2008, 4.2 million (28.5%) people with diabetes aged 40 years or older had diabetic retinopathy, and of these, 655000 (4.4% of those with diabetes) had advanced diabetic retinopathy that could lead to severe vision loss. These preoccupying figures also lead to increased costs for the community. In United states, it is estimated that the average medical costs of people with diabetes is 2.3 times higher than the expenditure for a patient without diabetes. This is likely to be due to the late diagnosis of diabetes-related conditions, either because of the lack of a specialized centre in an area nearby the patient or because of the specialized visit costs (Centers for Disease Control and Prevention, 2011).

The aim of this thesis is to harness and expand the capabilities of digital fundus cameras in order to create an inexpensive and high-throughput automated system to screen diabetic related retina disease (e.g. diabetic retinopathy and diabetic macular edema). A system able to correctly referring patients to an expert if a retina condition is detected would be of great help for diabetic patients, who can be screened remotely and inexpensively and ophthalmologists that would not otherwise be able to meet the health demands of an increasing diabetic population. The state of the art image processing, pattern recognition and machine learning algorithms have been researched and a series of new techniques have been developed and tested to reach this goal.

1.1 Aims and Objectives

The aim of this work is to develop a set of tools for the retina screening of diabetic patients uniquely with colour fundus images. During these years of research, four fixed objectives were given:

- **Objective 1**: develop an automated algorithm to assess the quality of fundus images.
- **Objective 2**: develop an algorithm to segment exudates in the context of other state of the art techniques.
- **Objective 3**: develop an algorithm to detect microaneurysms in the context of other state of the art techniques.
- **Objective 4**: evaluate the algorithms in the context of diabetic retinopathy screening.

In addition to the previous objectives, the research of other techniques in the context of diabetic retinopathy screening was encouraged. Therefore, after meeting the initial objectives two other routes were followed. A complete algorithm to detect diabetic retinopathy and diabetic macular edema (a dangerous retinopathy complication) from the lesions segmented was developed. Techniques to extend the functionality of standard fundus cameras with multiple images were investigated and two methods were developed: one to detect the swelling of the macula and one to increase the quality of fundus images after their acquisition.

1.2 Thesis Overview

This thesis is divided in 8 chapters. An outline of each one follows:

- **Chapter 2** introduces the basic biological informations required to fully comprehend the medical terms and techniques used throughout the thesis; it introduces the existing systems used to image the retina; it also presents the common image analysis techniques employed for the analysis of the eye fundus and the existing automated diagnosis approaches.
- **Chapter 3** presents an algorithm to automatically evaluate the quality of a fundus image.

- **Chapter 4** describes a rule-based method to segment exudates that does not require training on a labelled dataset. The HEI-MED dataset is introduced and employed to test the algorithm developed by comparing it to two other rule-based exudate segmentation algorithms.

- **Chapter 5** describes a microaneurysms localization method based on novel application of the Radon transform. The algorithm is tested on a public dataset which allows the comparison with techniques developed by other researchers.

- **Chapter 6** leverages the output of the lesion segmentation techniques to generate two diagnosis, one for diabetic retinopathy and another for diabetic macular edema. The algorithms are trained and tested on different public datasets one of which comprises of 1200 images and it provides a good estimation of the effective performance in a real world screening environment.

- **Chapter 7** introduces a new technique to detect the swelling of the macula by employing multiple fundus images. The technique is validated with a ray-traced model of the retina/eye/camera and with patient data.

- **Chapter 8** presents a retina image enhancer algorithm based on a novel way to use fundus cameras. Multiple images captured at different angles are employed to remove artefacts and other unwanted exposure problems.

Chapter 2

Background and Significance

"Les grandes personnes ne comprennent jamais rien toutes seules, et c'est fatigant, pour les enfants, de toujours et toujours leur donner des explications."
- Le Petit Prince, Antoine de Saint Exupéry

This chapter introduces the medical aspects and state of the art image analysis techniques required to fully comprehend the significance of fundus image analysis in diabetic patients. First, a general overview of the eye and the retina is given. Then, the morphology of different retina abnormalities is described. Particular attention is given to the lesions directly related to retinal diseases typical of diabetic patients. Also, a description of various types of retinopathy and related diseases are given. An overview of the public fundus image databases available at the time of writing follows. Of equal importance, the techniques used to evaluate, compare and contrast the results of the algorithms are described. Afterwards, a summary of the existing state of the art algorithms for the analysis of the morphology of the retina is given for the following aspects: normalization, vessel segmentation, optic disc / fovea localization and quality assessment. Additionally, comparisons of existing microaneurysms and exudates segmentation techniques are presented. Finally, two aspects of retina screening in diabetic patients are described: the various manual protocols available and current completely automated techniques for retina screening.

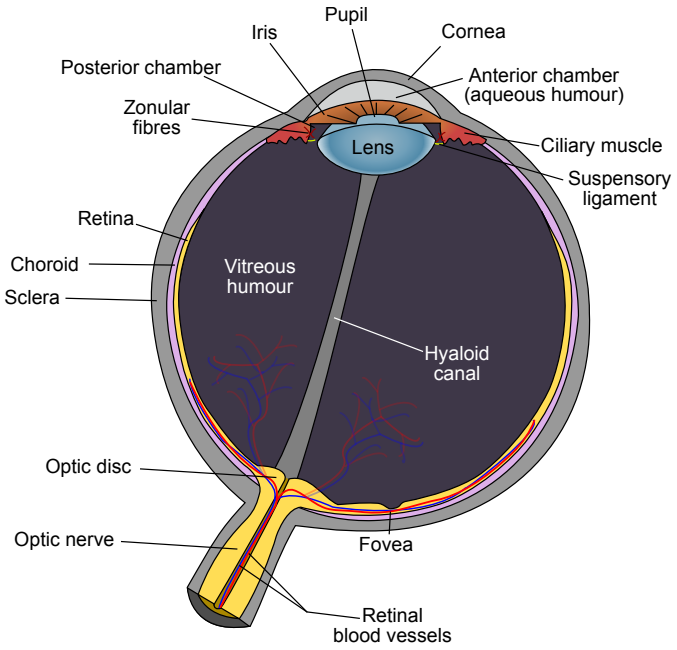


FIGURE 2.1: Global eye's anatomy.

2.1 the Eye and Retina

2.1.1 Eye's Anatomy

Vision is arguably the most used of the five senses in the human body. We rely on our eyes to provide most of the information we perceive about the world, so much so, that a significant portion of the brain is devoted entirely to visual processing. The eye is often compared to a camera because of the way it processes light into information understandable by the brain. Both have lenses to focus the incoming light. A camera uses the film to create a picture, whereas the eye uses a specialised layer of cells, called the retina, to produce an image. However the similarity stops here. The eye's ability of focus on a wide range of objects having different sizes, luminosity and contrast at a high speed are more powerful than those of current cameras.

Figure 2.1 shows a schematic view of the anatomy of eye. Light reaches the eye by first passing through the cornea which filters it, and begins focusing the image. The anterior chamber contains a viscous substance called aqueous humour, that keeps the front of the eye firm and slightly curved. Light travels through the pupil, which compensates for changing light conditions by contracting or relaxing. The muscles responsible for these movements are in the iris. Subsequently, the lens squeezes or stretches to focus the rays of light on the retina. The interior surface of the eye, opposite the lens, is called the fundus (Cassin and Solomon, 1990).

The retina is a multi-layered sensory tissue that lies on the back of the eye. It contains millions of photoreceptors that capture light rays and convert them into electrical impulses. These impulses travel along the optic nerve to the brain where they are converted into images. There are two types of photoreceptors in the retina: rods and cones, named after their shape. Rod cells are very sensitive to changes in contrast even at low light levels, hence able to detect movement, but they are imprecise and insensitive to colour. They are generally located in the periphery of the retina and used for scotopic vision (night vision). Cones, on the other hand, are high precision cells capable of detecting colours. They are mainly concentrated in the macula, the area responsible for photopic vision (day vision). The very central portion of the macula is called the fovea, which is where the human eye is able to best distinguish visual details. Whereas loss of peripheral vision may go unnoticed for some time, damage to the macula will result in loss of central vision, which has serious effects as illustrated in figure 2.8 (Wyszecki and Stiles, 1982). All the photoreceptors are connected to the brain through a dense network of roughly 1.2 million of nerves (Jonas et al., 1992). All nerves leave the eye in a unique bundle in the optic nerve. In the retina there is a blind spot which is a result of the absence of retina photoreceptors where the optic nerve leaves the eye.

Many retinal blood vessels supply the nutrients (oxygen and other components) to the inner and outer layer of the retina. The inner layer accounts for a smaller portion of the vessels ($\sim$35%), which are visible from the vitreous humour in common fundus images. The vessels in the outer layer are the source of $\sim$65% of nutrients for the retina, and they are rarely visible in fundus images since they are situated in the choroid (situated at the back of the retina). Fig. 2.2 shows the retina layers pictured in a lithograph plate from Gray's anatomy book, originally published in 1918. Currently, retina and choroid are divided into the following layers (Abramoff et al., 2010a):

1. **Internal limiting membrane** or membrana limitans interna.
2. **Nerve fibre layer** contains the axons of the ganglion cells, that transmit the signals to the visual cortex of the brain. This layer is much thicker in younger retinas and it slowly thins the more it ages. As the blood vessels in the inner layer, this fibre is not present over the fovea in order to permit all the photons to reach the rods without obstacles.
3. **Ganglion cell layer** contains the body of the ganglion cells.
4. **Inner plexiform layer** contains the axons of the bipolar cells and the amacrine cells.
5. **Inner nuclear layer** contains the cell bodies of the bipolar and horizontal cells.

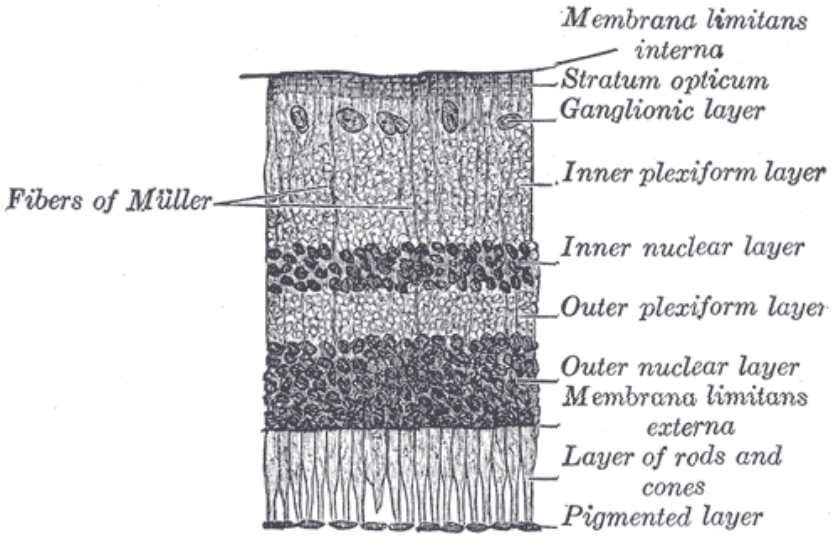


FIGURE 2.2: Retina Layers as depicted in Gray's anatomy book (Williams et al., 1989).

6. **Outer plexiform layer** contains the inner segments of the photoreceptor cells and the dendrites of the horizontal cells.
7. **Outer nuclear layer** contains the body of photoreceptor cells.
8. **External limiting membrane** or membrana limitans externa.
9. **Pigment epithelium**. It is the last layer of the retina. It is the one that gives a different colour (pigment) to each human retina.
10. **Bruch's membrane**. First layer of the choroid.
11. **Capillary choroid** contains all the blood vessels that provide the primary source of retina nutrients.
12. **Choroid plexus**. Last layer of the choroid.

The *ganglion cells* are a type of neuron that receive the signal from the photoreceptors through *bipolar cells* and *amacrine cells* (two intermediate neurons). The *horizontal cells* are interconnecting neurons that combine the signals received from photoreceptors (Masland, 2001).

2.1.2 Retina Imaging Techniques

Ophthalmologists conventionally imaged the eye using ophthalmoscopes which allowed for the determination of the health of the retina and vitreous humour. There are two types of ophthalmoscopes: direct and indirect. The former is an instrument

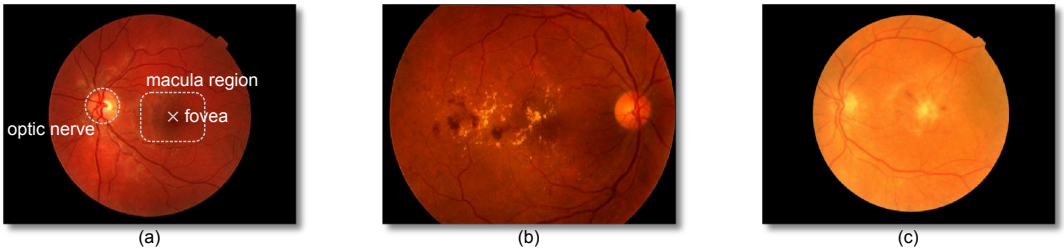


FIGURE 2.3: Examples of retina fundus images acquired with different cameras at 45°. (a) healthy patient; (b) retina showing signs of diabetic retinopathy (see Section 2.1.4.1); (c) retina showing signs of age related macular degeneration (see Section 2.1.4.3).

about the size of a small flashlight with several lenses that can magnify up to about 15 times, the latter is mounted on a headband and it provides a wider view of the fundus of the eye (Chernecky and Berger, 2004).

Relatively recently, fundus cameras were successfully introduced as an imaging technique (Hutchinson et al., 2000; Lin et al., 2002). A fundus camera provides an upright, magnified view of the fundus of the interior surface of the eye: the retina, optic disc, macula, and posterior pole (i.e. the fundus). A typical camera views 30 to 50 degrees of retinal area, with a magnification of 2.5x, and allows some modification of this relationship through zoom or auxiliary lenses from 15 degrees which provides 5x magnification to 140 degrees with a wide angle lens which minimises the image by half (Saine, 2006). Fig. 2.3 shows some examples of retina fundus images.

Currently, the preferred way to detect diseases like diabetic retinopathy is fundus camera imaging as illustrated by Luzio et al. (2004). Additionally, when digital photography is employed, the image may be enhanced, stored and retrieved more easily than film, and images may be transferred electronically to other sites for a trained optical technician or retinal specialist to diagnose diseases with the patient at a remote location. This is the modality used to acquire and analyse the vast majority of the images presented in this thesis.

Apart from colour digital fundus cameras, Abramoff et al. (2010a) add the following imaging modalities to a broader category of fundus imaging:

- **Stereo fundus photography**: at the same time two or more view angles of the fundus are acquired by this instrument. This allows the perception of the depth by the ophthalmologist.
- **Hyperspectral imaging**: it is a fundus camera that does not employ the visible light only, but can select specific wavelength bands. This allows particular applications such as *oximetry*, the quantification of oxygen levels in the bloodstream.

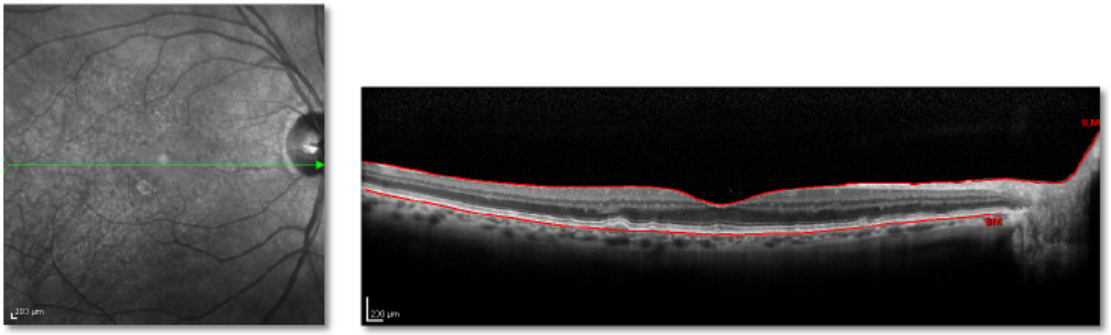


FIGURE 2.4: Example of a OCT slice acquired with a Heidelberg Spectralis (on the right). On the left the corresponding fundus image in the infrared. The green arrow shows the origin of the slice.

- **Fluorescein angiography (FA)**: is a fundus image of the photons emitted by a contrast agent injected in the patient's blood stream. Fluoroscein or indocyanine green fluorophore are the agents typically used.
- **Scanning laser ophthalmoscope (SLO)**: An instrument that uses low powered lasers to image the retina or choroid. It uses a very narrow moving beam of light which can bypass most ocular media opacities (i.e. corneal scars, cataracts, vitreous hemorrhage) to reach the surface of the retina and record its surface detail. With SLO, the optics of the eye serves as the objective lens. *Confocal SLO* is SLO equipped with a confocal aperture. *Adaptive optics SLO* optically corrects the laser reflections by modelling the aberrations in its wavefront.

The other imaging technique that is becoming increasingly important is Optical coherence tomography (OCT). OCT is a non-destructive imaging technique that uses interferometry techniques to measure the time of flight of the light backscattering through the retina. By rapidly scanning the eye, it can acquire an in vivo representation of the anatomic layers within the retina. Because of that it can be used to diagnose diseases such as DME, AMD and Glaucoma with generally a greater precision than with a simple fundus image (Walsh et al., 2010). Fig. 2.4 shows an example of a "retina slice" that can be acquired with these instruments.

However, DR cannot be directly diagnosed because the vessels and many other key features of the retina are invisible in this modality (even if it is possible to algorithmically infer the location of the vasculature by employing the visible shadows as shown by Niemeijer et al. (2008)). Other drawbacks of this modality are: the steeper learning curve to use the instrument than a colour fundus camera, the greater acquisition time required to acquire a field of view (FOV) comparable to a fundus camera and the substantially higher cost.

2.1.3 Retina Lesions and Abnormalities

2.1.3.1 Microaneurysms

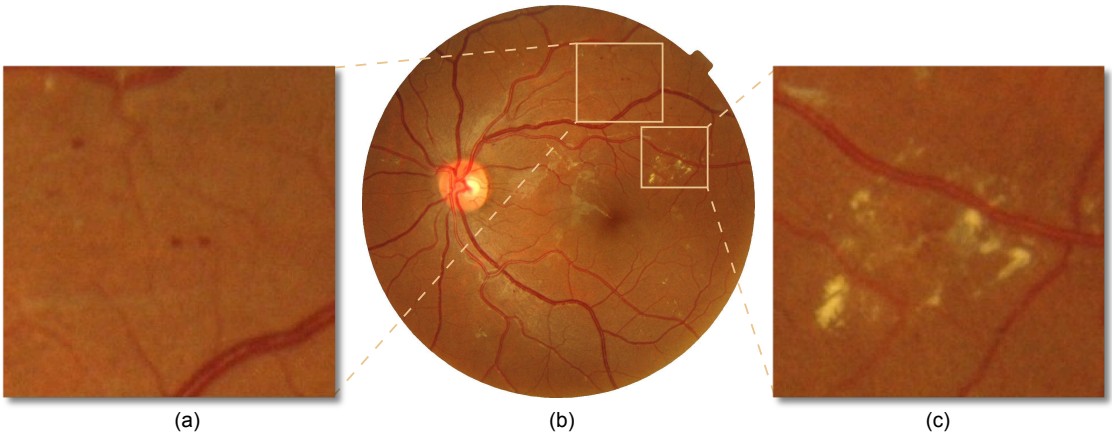


FIGURE 2.5: (a) Detail showing microaneurysms; (b) Fundus image; (c) Detail showing exudates.

Microaneurysms (MAs) are dilated, aneurismal retina vessels that appear as small red dots in colour retina fundus images. These lesions can leak fluid and blood into the retina, leading to vision threatening exudates, macular edema and hemorrhages. These MAs are the primary target lesions for laser treatment of diabetic retinopathy or macular edema. Fig. 2.5(a) shows an example of such lesions.

2.1.3.2 Hemorrhages

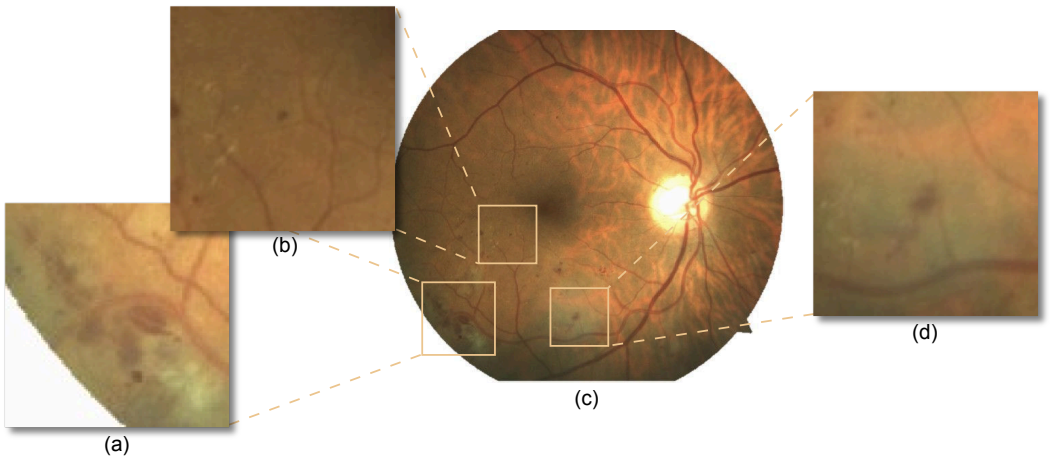


FIGURE 2.6: (a) Detail showing large superficial hemorrhages; (b) Fundus image; (c) Detail showing a dot hemorrhage; (d) Detail showing a blot hemorrhage.

Retinal hemorrhages (HMA) are loss of blood from the vasculature. They appear as red structures of variable shape in the fundus. Their shape can be correlated with the depth in the retina. Hemorrhages close to the ganglionic layer have a flame shape and tend to disappear within a short period of time. Blot and dot hemorrhages have a round shape (similar to large MAs) and are located in the retina's inner nuclear and outer plexiform layers. Their configuration is due to intraretinal compression, restricting the hemorrhages within a specific location (Niki et al., 1984). They are more serious because connected with diabetic retinopathy. Fig. 2.6 shows some examples of such lesions. When hemorrhages occur in the vitreous humour, they are called vitreous hemorrhages (VHs) or preretinal hemorrhages (PRHs) if they occur exactly between vitreous humour and retina. VHs and PRHs often happen because of neovascularization, see Section 2.1.3.6.

2.1.3.3 Hard Exudates

Hard exudates (HE or *Intra-retinal lipid exudates*) are yellow deposits of lipid and protein within the sensory retina. Fig. 2.5(c) shows some of them. They are often due the leakage of lipid and proteins from MAs which have walls thinner than normal capillaries due to the increased pressure. As such, exudates are the hallmarks for the diagnosis of macular edema from single fundus images because indicate increased vessel permeability and presence of fluid (see Section 2.1.4.2).

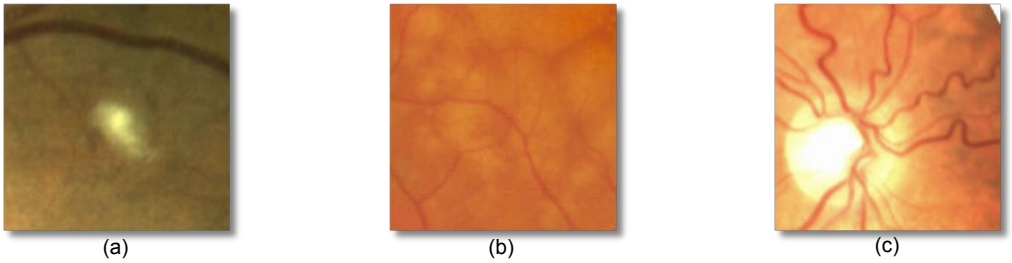


FIGURE 2.7: (a) Detail showing a cotton wool spot; (b) Detail of various drusen; (c) Detail some tortuous vessels.

2.1.3.4 Cotton Wool Spots

Cotton wool spots (CWS or *soft exudates*) are micro infarctions due to the lack of blood supply to the nerve fibre layer. They appear as yellowish/white structures with blurred edges. Cotton wool spots by themselves do not cause visual difficulties, but they are highly correlated with conditions that affect the retina circulation such as diabetic retinopathy. Fig. 2.7(a) shows an example of such lesions.

2.1.3.5 Drusen

Drusen are yellow deposits of extracellular material below the retina in the Bruch's membrane. They are a sign of macular degeneration (see Section 2.1.4.3). Since they are below the retina, they are sometimes not obvious in a fundus image, especially when only few of them are present. Fig. 2.7(b) shows an example.

2.1.3.6 Vascular Abnormalities

Hypertensions can also lead to other types of abnormalities in the retina vasculature. These abnormalities can be classified in various groups as reported by Grisan (2005). When the blood pressure increases, vessels may increase in length and thicken their walls resulting in some *vessel tortuosity* as shown in Fig. 2.7(c). Another effect is the *arteriolar narrowing* which is probably the earliest effect. Whenever veins have a diameter that do not decrease monotonically with distance from the macula, *venous beading* (VB) occurs. Gregson et al. (1995) note that beaded veins exhibit periodic changes in diameter along their lengths, somewhat like sausage links. *Bifurcation and crossing abnormalities* are other changes of the vascular path due to the abnormal pressure difference between veins and arteries. Other small abnormalities in the retina are called *intraretinal microvascular abnormalities* (IRMA).

Neovascularization is another vascular abnormality due to the lack of oxigen to an area of the retina. When this happens a signal protein called *vascular endothelial growth factor* (VEGF) is released and new vessels rapidly grow. These vessels are weaker than normal and are much more likely to create hemorraghes or to leak liquid or proteins. Often these vessels grow from the retina into the vitreous humour and upon their rupture, they can generate VHs and PRHs. Neovascularization can stem from the disc (NVD), the iris (NVI or rubeosis irides) or elsewhere (NVE).

2.1.4 Retinal and Related Diabetic Diseases

Diabetes mellitus (DM) is a chronic, systemic, life-threatening disease characterised by disordered metabolism and abnormally high blood sugar (hyperglycaemia) resulting from low levels of the hormone insulin with or without abnormal resistance to insulin's effects (Tierney et al., 2002). DM has many complications that can affect the eyes and nervous system, as well as the heart, kidneys and other organs. Diabetes is the leading cause of new cases of blindness among adults aged between 20 and 74 years. The Centers for Disease Control and Prevention estimates that 25.8 million people currently

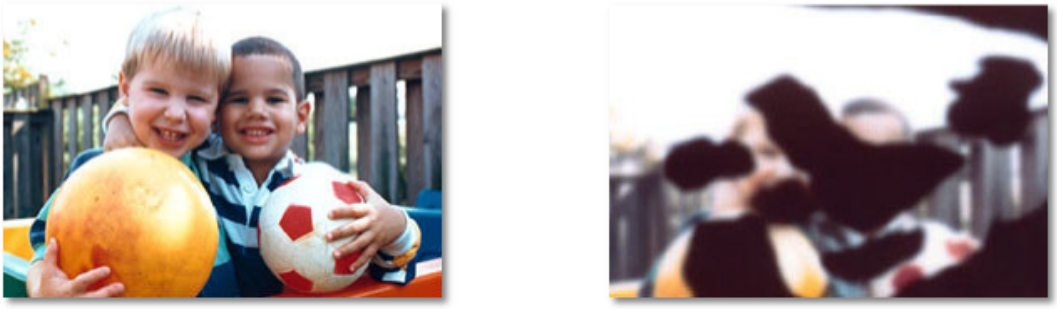


FIGURE 2.8: Loss of central vision due to diabetic retinopathy. On the left the vision of a healthy patient. On the right the vision of a patient with the retina damaged in the macula region. Image courtesy of the National Eye Institute.

have DM in the United States alone. The World Diabetes Foundation estimates that over 439 million people will have diabetes worldwide by 2030 (Centers for Disease Control and Prevention, 2011). Diabetic retinopathy and diabetic macular edema are the two retina diseases directly caused by diabetes.

2.1.4.1 Diabetic Retinopathy

Diabetic retinopathy (DR) is a vascular complication of DM which causes damages to the retina which leads to serious vision loss if not treated promptly. People with diabetes are 25 times more likely to develop blindness than individuals without diabetes. For any type of diabetes, the prevalence of diabetic retinopathy in people more than 40 years of age was reported to be 40.3% (Baker et al., 2008). MAs are a common and often early manifestation of DR. As such, the MA detector is an attractive candidate for an automatic screening system able to detect early findings of DR.

DR is generally graded in four subsequent stages (Joussen et al., 2007, chap. 19):

- **Mild Nonproliferative Retinopathy (Mild NPDR)**: At this earliest stage, only microaneurysms occur and possibly hemorrhage or hard exudates.
- **Moderate Nonproliferative Retinopathy (Moderate NPDR)**: As the disease progresses, some blood vessels that nourish the retina are blocked and consequently cotton wool spots start to appear. Also, initial vascular abnormalities such as tortuous vessels or IRMAs can be found.
- **Severe Nonproliferative Retinopathy (Severe NPDR)**: Many more blood vessels are blocked, depriving several areas of the retina with their blood supply. These areas of the retina send signals by the means of VEGF proteins and neovascularization starts.

- **Proliferative Retinopathy (PDR)**: At this advanced stage, the neovascularization start to reach the vitreous humour, the clear, vitreous gel that fills the inside of the eye. By themselves, these blood vessels do not cause symptoms or vision loss. However, they have thin, fragile walls. If they leak blood, vision loss and even blindness can result. Depending on the severity of the leakage, ophthalmologists grade the disease on a scale from 1 to 4. Neovascularization can remain relatively stable or it can grow rapidly, progression can be noted ophthalmoscopically over a period of weeks.

2.1.4.2 Diabetic Macular Edema

Diabetic macular edema (DME) is a complication of DR and is the most common cause of vision loss and blindness (Singer et al., 1992). DME is defined as swelling of the retina in diabetic patients due to leakage of fluid within the central macula from microaneurysms that form as the result of chronic damage due to elevated blood sugar levels. The presence of clinically significant DME is an important indication for the initiation of laser treatment. One particular characteristic of DME is thickening of the retina, which cannot be directly quantified from a single 2-D image because of the lack of depth information. Instead, ophthalmologists can infer the presence of the fluid that causes the retina thickening from photos, by the presence of accompanying lipid deposits called exudates. They appear as bright structures with well defined edges and variable shapes.

According to the Early Treatment Diabetic Retinopathy Study Research Group (ETDRS), the patient's retina should be immediately treated when DME is graded as *clinically significant*. This happens if any of the following points is satisfied (Joussen et al., 2007, chap. 19.3):

- There is a significant retinal thickening within 500 μm distance from the centre of the macula.
- There is exudation within 500 μm distance from the centre of the macula with retinal thickening in the bordering retina.
- There is a retinal thickened area by the size of at least one papilla diameter within the distance of one papilla diameter from the centre of the macula. The papilla roughly corresponds to the visible area covered by the ON in an image of the retina fundus.

2.1.4.3 Age-related Macular Degeneration

Another retinal disease that has similar effect to diabetic retinopathy but other causes is age-related macular degeneration (AMD). AMD refers to a degenerative condition affecting the macular or central area of the retina. The main characteristic of early AMD is the presence of soft drusen. While their presence does not necessarily indicate that a person has macular degeneration, there has been a strong correlation between the two and the presence of drusen substantially raises a person's risk to develop macular degeneration. Late AMD includes 2 distinct forms, dry AMD or wet. "Dry" AMD does not involve any leakage of blood or serum. Loss of vision may still occur. Patients with this "dry" form may have good central vision but substantial functional limitations including: fluctuating vision, difficulty reading because of their limited area of central vision and limited vision at night or under conditions of reduced illumination. In the "wet" type of macular degeneration, abnormal blood vessels grow under the retina and macula. In this case the patient will lose at least part of their central vision (Parmet, 2001).

2.1.4.4 Glaucoma

Glaucoma is not a retinopathy but a neuropathy. Nevertheless, it affects the retina by damaging ganglion cells and their axons. The disease is characterised by an increased pressure on the ON which slowly affects it, resulting in a peripheral visual field loss. Because of the peripheral damages, it is often unnoticed and it has become the second[1] leading cause of blindness in the world (Resnikoff et al., 2004).

Glaucoma can be diagnosed by analysing the 3-D shape of the optic nerve head. The ratio between the total diameter of the ON (*ON cup*) and the diameter of the area occupied by the optic nerve fibres is defined as cup-to-disc ratio, and it is an important indicator to evaluate the presence of the disease. In their thorough review, Abramoff et al. (2010a) present various methods to accomplish this task with fundus and OCT images.

[1] Note that the leading cause is cataract, which affect the cornea and is not a retina disease.

2.2 Retina Morphology Analysis in Fundus Images

2.2.1 Public Databases and Algorithms Evaluation

The design, implementation and testing of algorithms able to analyse the retina morphology or to diagnose a retina condition requires annotated data. In the medical imaging field, there are some publicly available annotated datasets of retinal images which have different goals, characteristics, and levels of completeness. Their aims include vessel segmentation, DR diagnosis and microaneurysm localization. However, the images by themselves are not enough to make a dataset particularly useful for algorithms development. The key aspect is the *ground truth* (GT) data which provides the gold standard against which the algorithms can be trained and tested. Whenever a single common dataset (with a common GT) has been employed by different research groups the advantages and disadvantages of different proposed methods were easily measurable and comparable as shown in Niemeijer et al. (2010). Table 2.1 shows a comparison of the retina datasets most commonly used based on their GT and number of images.

Dataset	Ves.	ON	MA	HE	OL	DR	ME	n° exp.	n° imgs.
STARE (Hoover et al., 2003)*	✓								81
DRIVE (Niemeijer et al., 2004)	✓							1	40
ARIA (ARIA, 2006)†	✓	✓							212
DIARETDB1 (Kauppi et al., 2007)			✓	✓	✓			4	89
MESSIDOR (Messidor, 2010)						✓	✓	1	1200
ROC (Niemeijer et al., 2010)			✓					4	100
HEI-MED (Giancardo et al., 2011c)		✓	✓	✓	✓		✓	1	169

∗ The GT of the STARE database is available upon request.
† The GT of the ARIA database is not available for all the images. The n° exp. varies between images.
Ves.: vessel segmentation GT; ON: optical nerve location GT; MA: microaneurysm segmentation GT; HE: hard exudates segmentation; OL: segmentation of other lesions GT; n° exp.: number of experts that generated the GT (if there are multiple experts for a dataset but the GT for each image comes from a single expert, this number is still 1); n° imgs.: number of images in the dataset.

TABLE 2.1: Public Fundus Images Datasets

Depending on the type of algorithm tested different evaluation strategies need to be adopted. A diagnostic system that outputs a binary healthy/unhealthy decision about an image can be evaluated against a GT (or *gold standard*) with *sensitivity* and *specificity*. The sensitivity is a ratio measured as the number of images classified as positive[2] that effectively were positives (true positives (TP)), divided by the "real" positives (i.e. false negatives (FN) plus TP). The specificity is a ratio measured as the number of images classified as negative that effectively were negative (true negative (TN)), divided by the "real" negatives (i.e. false positives (FP) plus TN). Both measures range between 0 and 1. This type of clear cut decisions are also defined as *categorical*. Another metric to

[2]Throughout this thesis we consider "positives" the samples that show a disease.

evaluate the categorical concordance between an algorithm and a gold standard is the Kappa value (k), which is widely used in the literature. The equation for k is:

$$Pr(a) = \frac{(TP + TN)}{n^{\circ}examples} \tag{2.1a}$$

$$Pr(e) = \frac{TP + FN}{n^{\circ}examples} \times \frac{TP + FP}{n^{\circ}examples} + \frac{FP + TN}{n^{\circ}examples} \times \frac{FN + TN}{n^{\circ}examples} \tag{2.1b}$$

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{2.1c}$$

Gwet (2002) underlines some pitfalls in k by showing its sensitivity to trait prevalence in the subject population and proposes $AC1$ as alternative:

$$P_{+} = \frac{(TP \times 2 + FN + FP)}{2 \times n^{\circ}examples} \tag{2.2a}$$

$$Pr(e_{AC1}) = 2 \times P_{+} \times (1 - P_{+}) \tag{2.2b}$$

$$AC1 = \frac{Pr(a) - Pr(e_{AC1})}{1 - Pr(e_{AC1})} \tag{2.2c}$$

Whenever the system tested still takes a binary decision but outputs a probability value rather than a categorical one, it can be better evaluated through a receiver operating characteristic (ROC) curve (Fawcett, 2004), a graphical curve which represents the *sensitivity* (on the vertical axis) and the complement to 1 of the *specificity* on the horizontal axis. Note that the ROC curve is not a single point (i.e. a single sensitivity/specificity pair), because the probability is thresholded at various points. A very compact representation of the ROC curve analysis is given by the *area under the ROC curve* (AUC), which is a single number between 0 and 1, where 1 denotes perfect performances and 0.5 a completely aleatory set of decisions. The AUC is also defined as "the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one" (Fawcett, 2006).

In some occasions the use of the TN is either not possible or not suggested. This is the case for the evaluation of lesion segmentation algorithms where the negative samples (that can be represented by pixels or other image areas) represent the vast majority of the image area. Hence, it is either not possible or misleading to calculate the specificity. A suited substitute is the *positive predictive value* (PPV), which is computed as $\frac{TP}{TP+FP}$. The free-response receiver operating characteristic (FROC) analysis (Metz, 1986) is the equivalent to the ROC curve with the PPV instead of 1-specificity. Some authors (such as Niemeijer et al. (2010)), plot directly the number of FP instead of the PPV. In this thesis both approaches are used.

Algorithm performance can also be measured with *accuracy*, a metric commonly employed in the field of science, engineering, industry and statistics especially for measurement systems. It evaluates the degree of closeness of measurements of a quantity to its actual (true) value (JCGM, 2004). The accuracy is computed as $\frac{TP+TN}{n^{\circ}examples}$.

2.2.2 Fundus Image Normalization

The first preprocessing step in many image processing applications is image normalization, and retina analysis is no exception. Some authors such as Sopharak et al. (2008) employ standard local or global histogram equalization techniques as described by Gonzales and Woods (2002). However, the majority takes approaches especially tailored for retina images which are also referred as *background subtraction* technique.

One of commonest is the one described by Spencer et al. (1996). They estimate the retina "background" with a large median filter on the green channel of the image, then they subtract the estimated background from the original image. Cree et al. (2005) refined the method by employing a target image to reduce the inter-patient colour variability. They assume that a background-less fundus image has colours normally distributed. Hence, two scalar values the mean (μ) and standard deviation (σ) across the entire image are sufficient to characterise it, and by taking a reference image and calculating these two parameters, it is possible to equalise the colours of the new image to the reference one in a more effective way than by simple histogram equalisation.

One shortcoming of the previous methods is the fact that the median filter background estimation might be biased by large foreground structures. Instead, Foracchia et al. (2005) identify the pixels belonging to the background by evaluating the mean and standard deviations on a local window. Once the pixels belonging to the background are identified, they are bilinearly interpolated to generate the complete background, which is then used to obtain the normalized image. Feng et al. (2007) adopt an approach based on the Contourlet, a multi-scale image analysis technique that has its roots in the Wavelets theory.

In Chapter 4 the fundus image normalization technique introduced in Giancardo et al. (2011d), is presented.

2.2.3 Vasculature Segmentation

The ability to discern vessels from other structures is a preprocessing step of great importance in many retinal imaging applications. The segmentation of the vasculature is

a key pre-processing step for various quality assessment, ON localization and microaneurysms detection algorithms. Additionally, the analysis of the vasculature width is an important clinical index in itself. Tramontan et al. (2011) describe a semi-automated system that measures the arteries to veins ratio (AVR) which seems to be a very early indication of the likelihood of developing retinopathy. Also, the tortuosity of the vessels is correlated with the abnormal blood pressure which appears to be an early index of various other retinopathies (Stanton et al., 1995; Grisan et al., 2008).

In Table 2.2, the performance of some of the leading algorithms on two public datasets are presented. The results are based on their respective publications. They have been divided in two classes: supervised and unsupervised. In the former, a classifier which requires a sound training set is generally required; in the latter, the algorithms are built with a series of image processing operators (rules), which might still require a training set for optimization of the parameters.

Type	Method	DRIVE		STARE	
		Accuracy	AUC	Accuracy	AUC
supervised	Staal et al. (2004)	0.9441	0.9520	-	-
	Niemeijer et al. (2004)	0.9417	0.9294	-	-
	Soares et al. (2006)	0.9466	0.9614	0.9480	0.9671
	Ricci and Perfetti (2007)	0.9595	0.9633	0.9646	0.9680
	Marin et al. (2011)	0.9452	0.9588	0.9526	0.9769
rule-based	Chaudhuri et al. (1989)	0.8773	0.7878	-	-
	Hoover et al. (2000)	-	-	0.9275	0.7590
	Jiang and Mojon (2003)	0.8911	0.9327	0.9009	0.9298
	Mendonca and Campilho (2006)	0.9463	-	0.9479	-
	Martinez-Perez et al. (2007)	0.9344	-	0.9410	-
	Cinsdikici and Aydin (2009)	0.9293	0.9407	-	-
	Zana and Klein (2001)	0.9377	0.8984	-	-

TABLE 2.2: Published vessel segmentation scores on public datasets

2.2.4 Optic Disc - Fovea localization

Together with the vasculature, the optic disc and fovea are the other two anatomic landmarks in the retina. In the context of abnormal retina detection the location and structure of the ON is important for at least two aspects: it is a confounding factor in the context of bright lesions segmentation (therefore it needs to be detected) and its morphology can be directly used for the detection of glaucoma (see Section 2.1.4.4). On the other hand, the location of the fovea can be used to judge the severity of an abnormality detected on the retina. As described in Section 2.4.1, lesions which are closer to the fovea (where our central vision resides) are more dangerous in terms of vision loss and potential vision loss.

The majority of algorithm employs the vascular directionality for ON localization (Foracchia et al., 2004; Tobin et al., 2007; Niemeijer et al., 2007a; Fleming et al., 2007; Niemeijer et al., 2009b; Hoover and Goldbaum, 2003). This is possible because all the vessels "enter" the retina from the ON. All the authors cited indicate a method for the localization of the fovea together with the ON.

Aquino et al. (2010) do not employ the vasculature, they propose a voting scheme based on three "simple" ON localization methods. These methods are based on local window pixel statics and are called: Maximum Difference, Maximum Variance and Low-Pass Filter Method. Although they do not present a technique for the localization of the macula, their method includes the segmentation of the ON boundary.

Karnowski et al. (2009) have also presented a hybrid ON localization technique based on vascular directionality and pure pixel classification using PCA and LDA. They found empirically that when the two approaches output a similar result it is extremely likely that the location of the ON is correct.

2.2.5 Quality Assessment

The computerized evaluation of image quality is a problem not only in the field of medical imaging but in many other image processing systems, such as image acquisition, compression, restoration and enhancement. However in fundus images the notion of quality is not in the image per se, but rather in the ability of an expert (human or automatic) to correctly diagnose the patient condition from the image. The fundus quality assessment methods can be grouped in three different categories depending on the technique used: histogram based methods, retina morphology methods and "bag-of-words" methods.

In histogram based methods, the histogram of certain features is calculated on a set of good quality images. Then the similarity with the reference histogram is used to evaluate the quality. Lee and Wang (1999) were the first authors that have explicitly addressed the problem of automatic detection of fundus image quality. Their approach starts from a pure signal processing perspective by employing the global histogram of the image intensities approximated by a Gaussian distribution. Lalonde et al. (2001) extended the approach of Lee and Wang using two different sets of features: the distribution of the edge magnitudes in the image and the local distribution of the pixel intensity. Rather than strictly correlating image quality with noise, their concept of quality depends on the experts' ability to perform a correct diagnosis.

Retina morphology methods employ features that are unique to fundus images to evaluate their quality. Usher et al. (2003) employ a vessel segmentation algorithm to estimate the image blurring. The area of the detected vessels is measured and directly correlated with image quality by the means of a hard threshold. Fleming et al. (2006a) presented a method that separately evaluates image clarity and field definition, which are finally combined to generate a global quality value. The image clarity was assessed taking into consideration the vessel area in the macula region, the field definition the relative position of fovea, ON and the length of the main vessel arcades.

In "bag-of-words" the authors borrow a method typically used for document classification. They employ a pattern recognition classifier (such as Support Vector Machine, Naive Bayes, etc.) to classify the occurrence of some common "words". These words are automatically generated from the raw features in test set with a unsupervised clustering algorithm (such as k-means). During the classification phase, the new raw features are associated to the known words based on their distance. Niemeijer et al. (2006) employed this approach with two sets of raw features to represent image quality: colour and second order image structure invariants (ISI). Colour is measured through the normalised histograms of the RGB planes, with 5 bins per plane. ISI are proposed by ter Haar Romeny (2003) who employed filterbanks to generate features invariant to rotation, position or scale. Recently, Paulus et al. (2010) employed a different set of features: the pixel grey levels and the Haralick texture metrics.

Chapter 3 describes the method proposed by Giancardo et al. (2010b).

2.3 Automatic Lesion Analysis

2.3.1 Microaneurysms Segmentation

In the early literature (Baudoin et al., 1984; Spencer et al., 1996; Frame et al., 1998), algorithms were developed to detect MAs (and other small round hemorrhages) in fluorescein angiograms (FAs). They employed a series of morphological operations which remove the vasculature, leaving the other small structures representing the MAs. These approaches perform well on FAs, where an intravenous contrast agent is injected into the subject; therefore the contrast between vessels/lesions and background is much greater than that of colour fundus images making the development of the algorithm straightforward. Consequently, the performance of these algorithms are not satisfactory on colour fundus images. In addition, the injection of the FA contrast agent is not a risk-free process and requires more highly trained personnel to be delivered than a

simple fundus photograph, hence it is not an optimal approach for broad-based DR screening (Lipson and Yannuzzi, 1989).

Niemeijer et al. (2010) announced the Retinal Online Challenge (ROC). The aim was to focus the efforts of the research community towards the creation of algorithms for the detection of MAs on colour fundus images, by evaluating their performance on a common dataset. Two sets of 50 images captured by different cameras that imaged the same area of the retina (45 degrees) were released to the research community together with a common evaluation modality. This allowed a fair comparison between algorithms developed by different groups. So far, 10 groups have participated in the challenge as shown in Table 2.3.

Group	Global Score	Sensitivity at 0.5 FPs
DRSCREEN (Antal et al., 2010)	0.434	0.380
ISMV II (Chapter 5)	0.375	0.366
Niemeijer et al. (2005)	0.395	0.336
LaTIM (Quellec et al., 2008)	0.381	0.318
OKmedical (Zhang et al., 2009)	0.357	0.315
OKmedical II	0.369	0.297
Lazar and Hajdu (2011)	0.355	0.274
Fujita Lab (Mizutani et al., 2009)	0.310	0.259
GIB Valladolid (Sanchez et al., 2009b)	0.322	0.254
ISMV (Giancardo et al., 2010a)	0.256	0.204
IRIA-Group	0.264	0.192
Waikato RIG	0.206	0.184

the table shows the ROC results as of March 2011, sorted by the third column. The details of the algorithms without an explicit reference can be found in the technical reports of the ROC website http://roc.healthcare.uiowa.edu/.

TABLE 2.3: Retinopathy Online Challenge Comparison

Typical approaches can be divided in three parts: an initial candidates detection, the removal of candidates that lie on the vessel segmented and a final classification of true positive and false positives by a pattern classification strategy. Sanchez et al. (2009b) and the Waikato detector start with a background subtraction with a large median filter, then two different adaptive thresholding approaches select the candidates. The candidates which intersect the automatically segmented vasculature are removed and finally a set of features is calculated and employed to train the classifiers. Zhang et al. (2009) perform similar steps by exploiting the responses from Gaussian filterbanks trained for the different purposes. Quellec et al. (2008) adopt a template matching approach on sub-bands of the wavelet space of the image, where the sub-bands are chosen during the training phase with a genetic algorithm strategy. Again, the candidates intersecting the automatically segmented vasculature are removed to obtain the final detection. Mizutani et al. (2009) detect the initial candidates and remove the ones lying the vasculature with two modified double ring filters. A set of 12 features is extracted for each

candidate and a classifier is trained in order to obtain the final output. Lazar and Hajdu (2011); Giancardo et al. (2010a) are the only techniques that are entirely rule-based, i.e. they do not explicitly require a training phase on labelled images. The former is based on a cross-section rotating model, and the latter on an operator based on the local Radon transform. Two of the top performing techniques (Antal et al., 2010; Niemeijer et al., 2005) are based on a mixture of multiple microaneurysm detection algorithms that are statistically combined together. In Chapter 5, we present our technique that rank among the firsts even if its computational and training complexity is minimal.

2.3.2 Exudates Segmentation

The approaches to exudate segmentation presented in the literature can be roughly divided into four different categories. *Thresholding methods* base the exudate identification on a global or adaptive grey level analysis (Phillips et al., 1993; Sanchez et al., 2009a). *Region growing methods* segment the images using the spatial contiguity of grey levels; a standard region growing approach is used by Sinthanayothin et al. (2002), which is computationally expensive. Li and Chutatape (2004) reduce the computational issues by employing edge detection to limit the size of regions. *Morphology methods* employ greyscale morphological operators to identify all structures with predictable shapes (such as vessels). These structures are removed from the image so that exudates can be identified (Walter et al., 2002; Sopharak et al., 2008). *Classification methods* build a feature vector for each pixel or pixel cluster, which are then classified by employing a machine learning approach into exudates or not exudates (Gardner et al., 1996; Osareh et al., 2003; Garcia et al., 2009) or additional types of bright lesions (Niemeijer et al., 2007b; Fleming et al., 2010) (such as drusen and cotton wool spots), all of which require good ground-truth datasets.

Table 2.4 compares various exudate segmentations techniques by showing the results reported by the respective authors. Unfortunately, all of them seem to employ different datasets and different evaluation methods. This makes a direct comparison almost impossible, as is underlined by many of the authors themselves. In a recent work (Giancardo et al., 2011c), we contribute to a solution to this problem by making available a heterogeneous, ground-truthed dataset to the research community: the HEI-MED dataset. A comparison of some exudate segmentation techniques is presented in Chapter 4.

evaluation	Authors	Sens.	Spec.	PPV	Requires Manual Segmentation	Images Tested	Processing Time
area	Gardner et al. (1996)	0.93	0.93	n/a	Yes	301	< 1 min.
	evaluation on a 20x20 grid; partially tested on the training set.						
	Li and Chutatape (2004)	1.00	0.71	n/a	No	35	n/a
	evaluation on 10 areas.						
	Sinthanayothin et al. (2002)	0.88	0.99	n/a	No	35	n/a
	evaluation on a 10x10 grid.						
	Phillips et al. (1993)	0.87	n/a	n/a	No	14	3 min.
	evaluation is pixel/area based on 1 area manually selected						
pixel	Niemeijer et al. (2007b)	0.95	0.86	n/a	Yes	300	n/a
	bright lesions are classified per type.						
	Osareh et al. (2003)	0.93	0.94	n/a	Yes	67	11 min.
	Sopharak et al. (2008)	0.80	0.99	n/a	No	60	3 min.
	Walter et al. (2002)	0.92	n/a	0.92	No	30	n/a
lesion	Sanchez et al. (2009a)	0.90	n/a	0.96	No	106	n/a
	Garcia et al. (2009)	0.87	n/a	0.83	Yes	67	n/a
	Giancardo et al. (2011d)	0.82	n/a	0.5	No	169	1.9 sec.

TABLE 2.4: Published Results on Exudate Segmentations

2.4 Screening Diabetic Patients

2.4.1 Screening Protocols

In order to monitor with fundus images diabetic retinopathy for research or clinical purposes, a standard acquisition protocol and grading scale need to be defined. According to Joussen et al. (2007, chap. 19.01), ETDRS supplies an acquisition protocol/-grading scale that have the best validation to date (Early Treatment Diabetic Retinopathy Study Research Group, 1991). The grading scale divides DR in 13 levels, on the basis of the relative risk of retinopathy progression and visual loss. It spans between absence of retinopathy to sever vitreous hemorrhage. The grading is based on the classification of the lesion severity in certain areas of the retina. A set of standard photographs are used in order to help the objective judgement of the lesion severity.

The standard acquisition protocol requires stereoscopic fundus images at 30° (captured on film), which are graded using a stereo viewer on a retroilluminated white glass plate. For each retina seven different images are captured with the configuration shown in Fig. 2.9, which is referred as *seven fields configuration*.

ETDRS grades DR and DME separately. Table 2.5 shows the ETDRS grading for DR. The grading for DME is considerably simpler, it comprises of three levels: DME absent,

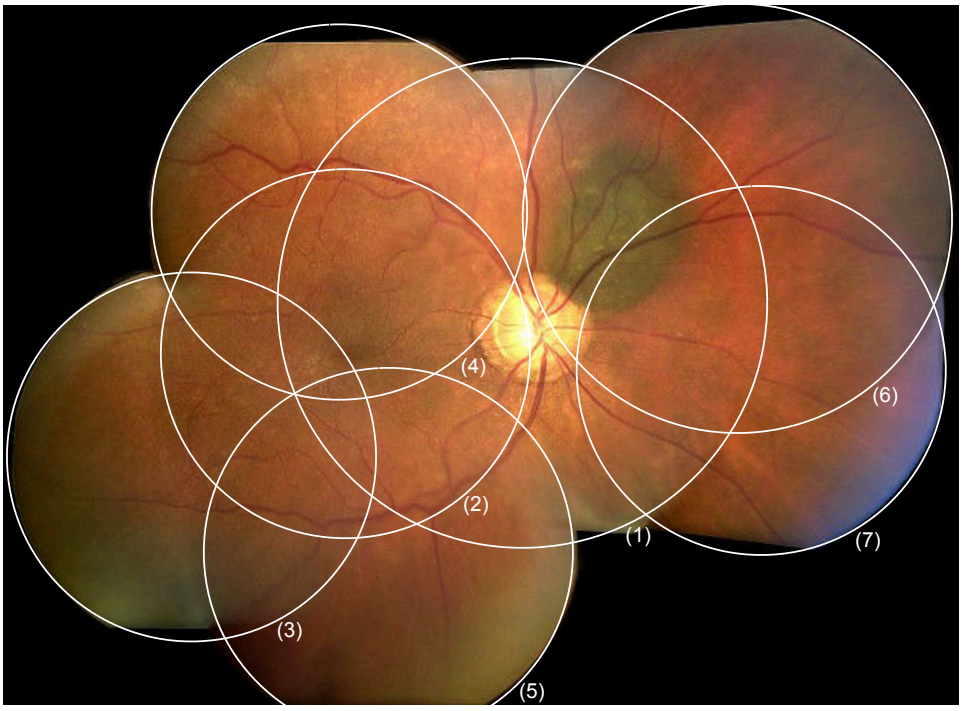


FIGURE 2.9: Example of registered fundus images acquired with the seven fields ETDRS specifications.

DME present or clinically significant DME (CSME). The ETDRS defines DME as the "thickening and/or hard exudate within 1 disk diameter of the centre of the macula". The properties for the CSME classification were described in Section 2.1.4.2.

In order to accommodate larger epidemiological studies with reading centres having a limited experience with fundus photography, the EURODIAB system was introduced (Aldington et al., 1995). The protocols involve two (monocular) fundus images per eye at 45°: one image centred on the macula and one centred on the optic disk. The DR grading has 6 levels only (from 0 to 5) and is based on the experience acquired by ETDRS. Table 2.5 also shows the direct correlation between ETDRS and EURODIAB grading.

Wilkinson et al. (2003) proposed a scale developed by a consensus panel of experts to provide a scale that can be used in a practical clinical settings. The *International Clinical Diabetic Retinopathy and Diabetic Macular Edema Disease Severity Scale* classifies DR in five stages and DME in four stages as shown in Table 2.6.

Williams et al. (2004) conducted a review to further simplify the acquisition protocol by employing a single fundus image. The authors concluded that although it cannot be a substitute for a comprehensive ophthalmic examination, "there is level I evidence that it can serve as a screening tool for diabetic retinopathy to identify patients with

ETDRS level	ETDRS severity	ETDRS definition	EURODIAB level
10	No retinopathy	Diabetic retinopathy absent	0
20	Very mild NPDR	Microaneurysms only	1
35	Mild NPDR	Hard exudates, cotton-wool spots, and/or mild retinal hemorrhages	1
43	Moderate NPDR	43A: retinal hemorrhages moderate ($>$ photograph 1a) in four quadrants or severe (photograph 2A) in one quadrant 43B: mild IRMA ($<$ photograph 8A) in one to three quadrants	2
47	Moderate NPDR	47A: both level 43 characteristics 47B: mild IRMA in four quadrants 47C: severe retinal hemorrhage in two to three quadrants 47D: venous beading in one quadrant	2
53A-D	Severe NPDR	53A: $\geq$ 2 level 47 characteristics 53B: severe retinal hemorrhages in four quadrants 53C: moderate to severe IRMA in at least one quadrant 53D: venous beading in at least 2 quadrants	3
53E	Very severe NPDR	$\geq$ 2 level 53A-D characteristics	3
61	Mild PDR	NVE $<$ 0.5 disk area in one or more quadrants	5
65	Moderate PDR	65A: NVE $\geq$ 0.5 disk area in one or more quadrants 65B: NVD $<$ photograph 10A ($<$ 0.25 - 0.33 disk area)	5
71,75	High-risk PDR	NVD $\geq$ photograph 10A, or NVD < photograph10A or NVE $\geq$ disk area plus VH or PRH, or VH or PRH obscuring $\geq$ 1 disk area	5
81,85	Advanced PDR	Fundus partially obscured by VH and either new vessels ungradable or retina detached at the centre of the macula	5
	(none)	EURODIAB only level: photocoagulated	4

TABLE 2.5: Comparison of ETDRS levels with EURODIAB levels for Diabetic Retinopathy. The reference images can be found in (ETDRS, 1991). The acronym definitions are in Section 2.1.3

retinopathy for referral for ophthalmic evaluation and management". This study opens the doors for a much simplified protocol for manual and automatic screening systems.

Understandably, many other screening protocols that employ fundus images exist. An example is the *Atherosclerosis Risk in Communities* (ARIC) developed to investigate factors associated with both atherosclerosis and incidence of clinical cardiovascular disease through the retina (Wong et al., 2001).

2.4.2 Automatic Screening Systems

The ability to detect lesions (or other abnormalities) automatically in fundus images naturally leads to the creation of systems which can generate a diagnosis without human intervention. The diagnosis can be used either as a suggestion for the human reader (Computer Assisted Diagnosis or CAD) or as screening system. The latter is

Diabetic Retinopathy	Severity level	Findings
	No apparent retinopathy	No abnormalities
	Mild non-proliferative DR	Microaneurysms only
	Moderate non-proliferative DR	More than just microaneurysms but less than severe non-proliferative diabetic retinopathy
	Severe non-proliferative DR	Any of the following: more than 20 HMA in each of 4 quadrants; definite VB in 2+ quadrants; prominent IRMA in 1+ quadrant and no signs of proliferative DR
	Proliferative DR	One or more of the following: neovascularization, vitreous/preretinal HMA
Diabetic Macular Edema	**Severity level**	**Findings**
	No apparent DME	No apparent retinal thickening or hard exudates in posterior pole
	Mild DME	some retinal thickening or hard exudates but distant from the centre of the macula
	Moderate DME	retinal thickening or hard exudates approaching (without involving) the centre of the macula
	Severe DME	retinal thickening or hard exudates involving the centre of the macula

TABLE 2.6: International Clinical Diabetic Retinopathy and Diabetic Macular Edema Disease Severity Scale (Wilkinson et al., 2003). The acronym definitions are in Section 2.1.3

the application that would bring the greatest benefits for patients with diabetes or pre-diabetes, especially if implemented as an easy to operate, low-cost system capillary installed in pharmacies and shopping malls. In fact, in the United States alone there are currently 7 million diabetic patients that remain undiagnosed, and this number is destined to rise given the increasing incidence of diabetes (Centers for Disease Control and Prevention, 2011). The recommended level of screening[3] is not currently being achieved because many patients do not have access to specialized ophthalmologic centres because of their location, insufficient local specialists for screening, or healthcare costs.

The first large study to test the feasibility of such system is the *EyeCheck project* (Abramoff and Suttorp-Schulten, 2005). The authors acquired two fundus images from each of the 17877 patients with diabetes who had not been previously diagnosed with DR. All images were classified by a retina expert with a simplified grading scale than the ones presented in Section 2.4.1. Each image is graded as having no DR or more than minimal DR, in order to classify the patients that need referral to an ophthalmologist or not. As suggested by Williams et al. (2004) only a single field fundus image is employed for each eye, which greatly simplify the acquisition process.

[3]The recommended levels of screening for a diabetic patient (without conclamated diabetic retinopathy) is generally considered to be once per year (Melville et al., 2000). However, there are some studies that challenges this time frame (Vijan et al., 2000).

Abramoff et al. (2010b) obtained an AUC of 0.839 for the original algorithm developed in Eyecheck project, an AUC of 0.821 for another algorithm developed in the ROC challenge and an AUC of 0.86 when the two techniques were combined. These results are obtained by uniquely employing dark lesions (MA and HE) detectors. Instead, Niemeijer et al. (2009a) obtained an AUC of 0.881 by fusing together dark and bright lesion detectors with other miscellaneous features. In this case the tests were run on a subset containing data from 15000 patients instead of 17877.

It should be highlighted that the results evaluation are based on *ground truth* (GT) data generated by experts that is not perfect by any means, either because of human error or the sub-optimality of the imaging process. Based on intra-readers agreement calculation performed on the diagnosis generated by three experts on a set of 500 exams, Abramoff et al. (2010b) calculates that an AUC of 0.86 is the theoretical expected maximum, as greater values would not indicate true performance increase. However, this calculation is applicable only on these particular exams, where there was an average agreement of $k = 0.44$ between the three experts. Philip et al. (2007) report an inter-reader agreement of $k = 0.91$ for the screening outcome on a set of 14406 images captured from 6722 patients. Without having access to the original data it is impossible to calculate the theoretical maximum AUC, however it would be higher than the one presented by Abramoff et al. (2010b). In their study Philip et al. (2007), do not evaluate their results by the means of a ROC curve, but rather single specificity/sensitivity pairs. In a patient based evaluation, they obtain a sensitivity/specificity of 0.905/0.674; in a image based evaluation, they report a score of 0.862/0.768.

Since 2005, the Hamilton Eye Institute (University of Tennessee) in collaboration with the Oak Ridge National Laboratory and the University of Burgundy has been designing and developing a HIPAA-compliant, teleophthalmology network for the screening of diabetic retinopathy and related conditions in under served regions of the Mid-South of the United States of America. As of 2011, five clinics are provided with Zeiss Visucam PRO fundus cameras appropriately modified to automatically send pictures and patients' metadata to our central server for a diagnosis. The current telemedicine network is described by Li et al. (2011). The work flow of this system is currently semi-automated, with the prospect of moving to a fully automated screening system upon the acceptance of ophthalmologists, patients and Federal Drug Administration (FDA) approval. Also, it is running some of the algorithms presented in this thesis in a real world setting which has been of great help during the development.

Chapter 3

Quality Assessment

"Very often do the captains of such ships take those absent-minded young philosophers to task, upbraiding them with not feeling sufficient "interest" in the voyage; half-hinting that they are so hopelessly lost to all honorable ambition, as that in their secret souls they would rather not see whales than otherwise. But all in vain; those young Platonists have a notion that their vision is imperfect; they are short-sighted; what use, then, to strain the visual nerve? They have left their opera-glasses at home."
- The Mast-Head, Moby Dick: or the Whale, Herman Melville

In a real world screening environment, there are many unpredicted aspects that can result in an ungradable image. For this reason an algorithm able to automatically judge the quality of the fundus image captured is a necessary preprocessing step for a correct diagnosis. Some methods that address this problem can be found in the literature but they have some drawbacks like accuracy or speed. In this chapter a method for automatically judging the image quality with a score from 0 to 1 is presented. To do so various descriptors are used, among them a new set of features independent of field of view or resolution to indirectly describe the morphology of the patient's vessels. A C++ implementation of the complete algorithm is developed and tests suggest that the image quality can be estimated in $\sim$2 seconds on a 3.4 GHz Intel Pentium 4 machine with 2 GB of RAM. Part of this work has been published in Giancardo et al. (2008, 2010b).

3.1 Introduction

Various systems for automatic or semi-automatic detection of retinopathy with fundus images have been developed. The results obtained are promising but the initial image quality is a limiting factor (Patton et al., 2006); this is especially true if the machine operator is not a trained photographer. Algorithms to correct the illumination or increase the vessel contrast exist (see Section 2.2.2), however they cannot restore an image beyond a certain level of quality degradation. On the other hand, a fast and accurate quality assessment algorithm would automatically warn the operator to re-take the fundus image, eliminating or reducing the need for correction algorithms. In addition, a quality metric permits the automatic submission of the best images, if many are available.

The measurement of a precise image quality index is not a straightforward task, mainly because quality is a subjective concept which varies even between experts, especially for images that are in the middle of the quality scale. In addition, image quality is dependent upon the type of diagnosis being made. For example, an image with dark regions might be considered of good quality for detecting glaucoma but of bad quality for detecting diabetic retinopathy. For this reason, we decided to define quality as the *"characteristics of an image that allow the retinopathy diagnosis by a human or software expert"*.

Fig. 3.1 shows some examples of macula centred fundus images whose quality is very likely to be judged as poor by many ophthalmologists. The reasons for this vary. They can be related to the camera settings like exposure or focal plane error (Fig. 3.1(a,e,f)), the camera condition like a dirty or shuttered lens (Fig. 3.1(d,h)), the movements of the patient which might blur the image (Fig. 3.1(c)) or if the patient is not in the field of view of the camera (Fig. 3.1(g)). We define an outlier as any image that is not a retina image which could be submitted to the screening system by mistake.

Existing algorithms to estimate the image quality are based on the length of visible vessels in the macula region (Fleming et al., 2006a), or edges and luminosity with respect to a reference image (Lee and Wang, 1999; Lalonde et al., 2001). Another method uses an unsupervised classifier that employs multi-scale filterbanks responses (Niemeijer et al., 2006). The shortcomings of these methods are either the fact that they do not take into account the natural variance encountered in retinal images or that they require a considerable time to produce a result. See Section 2.2.5 for more details.

Additionally, none of the algorithms in the literature that we surveyed generate a "quality measure". Authors tend to split the quality levels into distinct classes and to classify images in particular ones. This approach is not really flexible and is error prone.

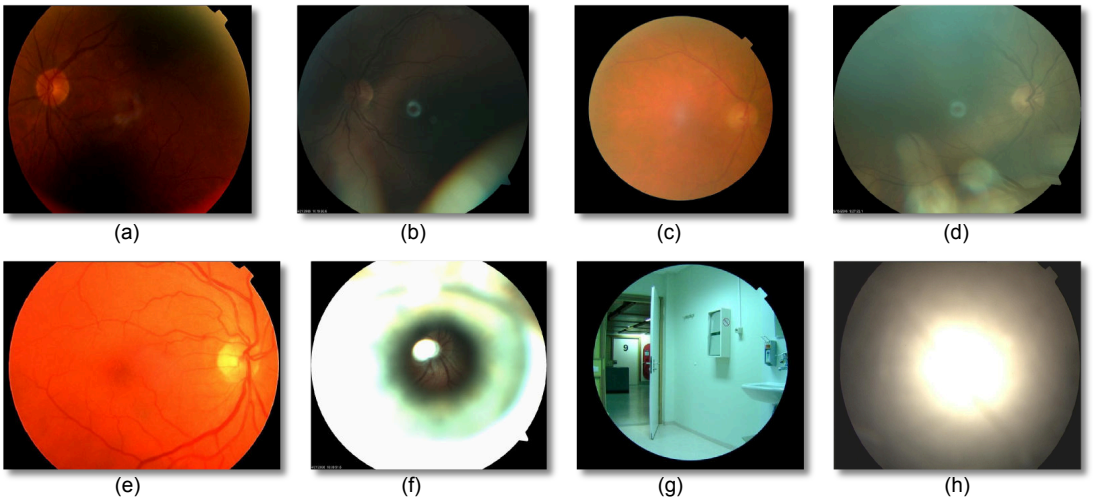


FIGURE 3.1: Examples of Poor Quality Fundus Images. (a) Underexposed; (b) eyelid artefacts; (c) blurred; (d) dirty lens; (e) overexposed; (f) out of focal plane; (g) no patient (outlier); (h) camera objective covered (outlier).

In fact human experts are likely to disagree if many categories of image quality are used. Therefore, a "quality measure" between 0 and 1 seems the most indicated for this classification problem.

Processing speed is another aspect to be taken into consideration. While algorithms to assess the disease state of the retina do not need to be particularly fast (within reason), the time response of the quality evaluation method is key towards the development of an automatic retinopathy screening system which allows real-time interactions with the photographer.

This chapter is structured as follows. Section 3.2 describes the proposed quality assessment technique based on the ELVD features. Section 3.3 describes the tests and results obtained. Section 3.4 concludes the chapter.

3.2 Methodology

The QA proposed aims to be: *accurate* in its QA of patients of different ethnicities, *robust* enough to be able to deal with the vast majority of the images that a fundus camera can produce (outliers included), *independent* of the camera used, *computationally inexpensive* so that it can produce a QA in a reasonable time and, finally it should produce a *quality index* from 0 to 1 which can be used as input for further processing.

Our approach is based on the hypothesis that a vessel segmentation algorithm's ability to detect the eye vasculature correctly is partly correlated to the overall quality of an

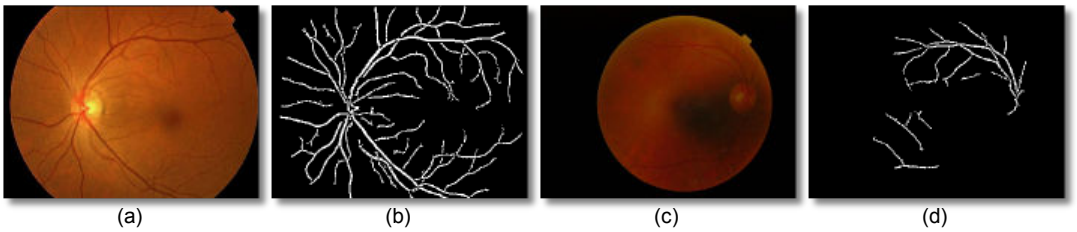


FIGURE 3.2: Comparison of the vessel segmentation by our implementation of Zana and Klein (2001) in a good and a poor quality fundus image. (a) Fundus image: good quality; (b) vessel segmentation of (a); (c) fundus image: bad quality; (d) vessel segmentation of (c).

image. Fig. 3.2 shows the output of the vessel segmentation algorithm in images with different quality. It is immediately evident that the low vessel density in the bottom part of the right image is due to an uneven illumination and possibly to some blurring. However, a global measure of the vessel area (or vessel density) is not enough to discriminate good from bad quality images. One reason is that a considerable quantity of vessels area is taken by the two arcades which are likely to be detected even in a poor quality image as in Usher et al. (2003). Another problem is that the illumination or blurring might be uneven, making only part of the vessels undetectable. The visible part can be enough to trick the QA. Finally, this type of measure does not take into account outliers or artefacts caused by smudges on the lens.

The algorithm presented is divided in three stages: Preprocessing, Features Extraction and Classification. An in depth illustration of the full technique follows in the next sections.

3.2.1 Preprocessing

3.2.1.1 FOV Mask Segmentation

In each digital fundus image acquired by modern fundus cameras there is a dark area surrounding the effective FOV showing the retina fundus defined as the *FOV mask*. Depending on the settings, each fundus camera has a FOV mask of different shape and size, so identifying which pixels actually belong to the retina is a necessary step. Some fundus cameras provide the FOV mask information. However, having the ability to automatically detect the FOV mask has some benefits. It improves the compatibility across fundus cameras because it does not need to be interfaced with any sort of proprietary format to access the FOV mask information. Also, if the QA is performed remotely, it reduces the quantity of information to be transmitted over the network.

Finally, some image archives use a variety of fundus cameras and the FOV mask is not known for each image.

The FOV mask segmentation is based on region growing (Gonzales and Woods, 2002). It starts by extracting the green channel of the RGB fundus image, which contains the most contrast between the physiological features in the retina (Teng et al., 2002), hence this channel best describes the boundary between background and foreground. It is also the channel that is typically used for vessel segmentation. Then, the image is scaled down to a height of 120 pixel (maintaining the aspect ratio), an empirically derived resolution which keeps the computational complexity as low as possible. Four seeds are placed on the four corners of the image with an offset equals to 4% of the width or height.

$$\begin{aligned} offset_w &= round(imageWidth \times 0.04) \\ offset_h &= round(imageHeight \times 0.04) \\ seed_{tl} &= [offset_w; offset_h] \\ seed_{tr} &= [imageWidth - offset_w; offset_h] \\ seed_{bl} &= [offset_w; imageHeight - offset_h] \\ seed_{br} &= [imageWidth - offset_w; imageHeight - offset_h] \end{aligned} \tag{3.1}$$

where $seed_{xy}$ is the location of a seed. The reason for the offsets is to avoid regions getting "trapped" by watermarks, ids, dates or other labels that generally appear on one of the corners of the image.

The region growing algorithm is started from the 4 seeds with the following criteria:

1. The absolute grey-level difference between any pixel to be connected and the mean value of the entire region must be lower than 10. This number is based on the results of various experiments.
2. To be included in one of the regions, the pixel must be 4-connected to at least one pixel in that region.
3. When no pixel satisfies the second criterion, the region growing process is stopped.

When four regions are segmented, the FOV mask is filled with negative pixels when it belongs to a region and positive otherwise. The process is completed scaling back the image to its original size. Even if this final step leads to a slight quality loss, the advantages in terms of computational time are worth the small imperfections at the edges of the FOV mask.

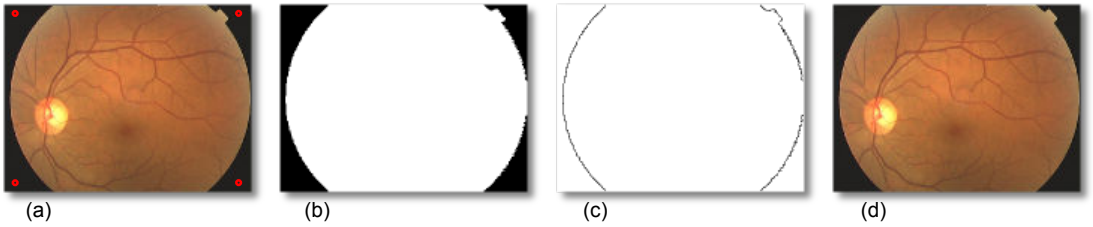


FIGURE 3.3: (a) Original image with the 4 seeds (in red) placed. (b) Mask segmentation results. (c) Points used for VFOV detection. (d) VFOV detected.

3.2.1.2 "Virtual" FOV Identification

During the acquisition of a macula centred image, the patient is asked to look at a fixed point visible at the back of the camera lens. In this way the macula is located at the centre of the image Field of View (FOV). Even if the area viewed by different cameras is standardized, various vendors crop some part of the fundus images that do not contain useful information for diagnosis purposes.

In order to develop an algorithm that runs independently from the lost information, the "Virtual" FOV (VFOV) is extracted. The VFOV consists of an ellipse that represents the contour of the fundus image as if it was not cropped. This measure allows a simplification of the algorithm at further stages and it is the key component that makes the method independent of the camera FOV and resolution.

The classical technique to fit a geometric primitive such as an ellipse to a set of points is the use of iterative methods like the Hough transform (Leavers, 1992) or RANSAC (Rosin, 1993). Iterative methods, however, require an unpredictable amount of computational time because the size of the image mask could vary. Instead, we employ the non-iterative least squares based algorithm presented by Halir and Flusser (2000) which is extremely computationally efficient and predictable.

The points to be fitted by the ellipse are calculated using simple morphological operations on the mask. The complete procedure follows.

$$\begin{aligned} \alpha &= erode(maskImage) \\ \gamma &= maskImage - \alpha \\ & fitEllipse(\gamma) \end{aligned} \tag{3.2}$$

The erosion is computed with a square structuring element of 5 pixels. The binary nature of the image in this step (Fig. 3.3.b) makes the erosion very computationally efficient.

3.2.1.3 Vessel Segmentation

The technique chosen to segment veins and arteries visible in fundus images is based on the mathematical morphology method introduced by Zana and Klein (2001). This algorithm proved to be effective in the telemedicine automatic retinopathy screening system currently developed in our group (see Section 2.4.2). Having multiple modules that share the same vessel segmentation algorithm is a benefit for the system as a whole to prevent redundant processing.

As shown in Section 2.2.3, the Zana & Klein algorithm are an example of a rule-based vascular segmentation method, and thus allows the usage of the algorithm without a training phase with manually labelled vessels. Our implementation is fast and efficient (<2 seconds as shown in Table 3.4) and it has been shown to work well even in infant fundus images where the choroidal vasculature poses a problem to many techniques (Heneghan et al., 2002).

This algorithm makes extensive use of morphological operations, the following acronyms are used:
erosion: $\epsilon_B(S)$
dilation: $\delta_B(S)$
opening: $\gamma_B(S) = \delta_B(\epsilon_B(S))$
closing: $\phi_B(S) = \epsilon_B(\delta_B(S))$
geodesic reconstruction (or opening): $\gamma^{rec}_{S_{marker}}(S_{mask})$
geodesic closing: $\phi^{rec}_{S_{marker}}(S_{mask}) = N_{max} - \gamma^{rec}_{N_{max}-S_{marker}}(N_{max} - S_{mask})$

where B is the structuring element and S is the image to which it is applied, S_{marker} is the marker, S_{mask} is the mask and S_{max} is the maximum possible value of the pixel. A presentation of these morphological operators can be found in Vincent (1993).

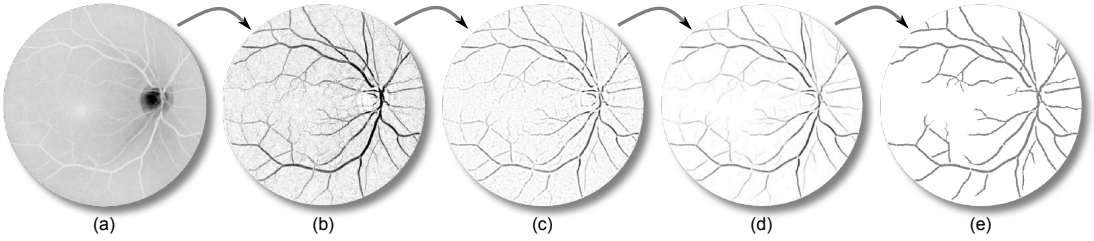


FIGURE 3.4: Vessel segmentation summary. (a) Initial image (green channel). (b) Image after Eq. 3.4. (c) Image after Gaussian and Laplacian filter. (d) Image after Eq. 3.7. (e) Final segmentation after binarisation and removal of small connected components. All images, apart from the first one, have been inverted to improve the visualisation.

The vessel segmentation starts by using the inverted green channel image already extracted by the mask segmentation. In fact, the blue channel appears to be very weak without much contrast with the vessels. On the other hand, the red band is usually too saturated since vessels and other retinal features reflect most of their signal in the red wavelength.

The initial noise is removed while preserving most of the capillaries on the original image S_0 as follows:

$$S_{op} = \gamma^{rec}_{S_0}(Max_{i=1...12}\{\gamma_{L_i}(S_0)\}) \tag{3.3}$$

where L_i is a linear structuring element 13 pixels long and 1 wide for a fundus image. For each i, the element is rotated of 15°. The authors specify that the original method is not robust for changes of scale. However, since we have an estimation of the VFOV, we are in a position to dynamically adapt the size of the structuring elements depending on the length of the axes in the VFOV.

Vessels can be considered as linear bright shapes identifiable as follows:

$$S_{sum} = \sum_{i=1}^{12}(S_{op} - \gamma_{L_i}(S_0)) \tag{3.4}$$

The previous operation (a sum of top hats) improves the contrast of the vessels but at the same time various unwanted structures will be highlighted as well. The authors evaluate the vessel curvature with a Gaussian filter (width=7px; $\sigma = 7/4$) and a Laplacian (size=3x3) obtaining the image S_{lap}. Then, by iterating the following operation the final result is obtained and the remaining noise patterns eliminated:

$$S_1 = \gamma^{rec}_{S_lap}(Max_{i=1...12}\{\gamma_{L_i}(S_lap)\}) \tag{3.5}$$

$$S_2 = \phi^{rec}_{S_1}(Min_{i=1...12}\{\phi_{L_i}(S_1)\}) \tag{3.6}$$

$$S_{res} = (Max_{i=1...12}\{\gamma^2_{L_i}(S_2)\} \geq 1) \tag{3.7}$$

As the last step of our implementation, we binarize the image and remove all the connected components that have an area smaller than 250 pixels. Once again this value is scaled depending on the VFOV detected. Fig. 3.4 shows a visual summary of the whole algorithm.

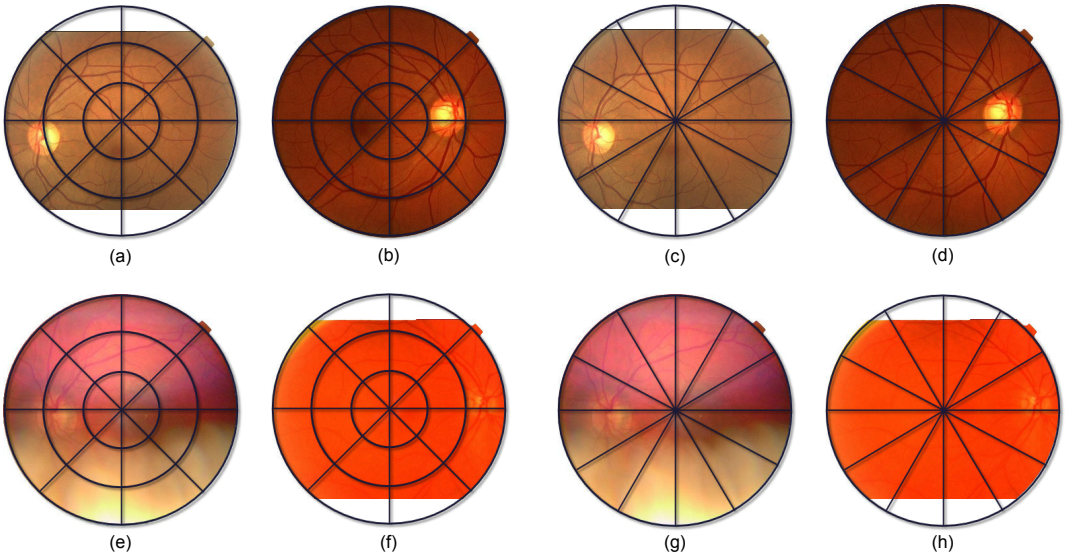


FIGURE 3.5: Elliptical local vessel density examples. Even and odd columns respectively contain left and right retina images. In top row good quality images are shown, in the bottom row bad quality ones. The 4 images on the left use ELVD with $\theta = 8$ and $r = 3$; the 4 images on the right are the same ones but the parameters for ELVD are $\theta = 12$ and $r = 1$.

3.2.2 Feature Extraction

3.2.2.1 Elliptical Local Vessel Density (ELVD)

By employing all information gathered in the preprocessing phase, we are able to extract a local measure of the vessel density which is camera independent and scale invariant. Other authors either measure a similar feature globally like Usher et al. (2003), or they use a computationally expensive method like Fleming et al. (2006a) whose approach requires a vessel segmentation, a template cross correlation and two different Hough transforms. Instead, we employ an "adaptable" polar coordinate system (θ, r) with the origin coincident with the origin of the VFOV. It is adaptable in the sense that its radius is not constant but it changes according to the shape of the ellipse. This allows the algorithm to deal with changes of scale not proportional between height and width.

The Elliptical Local Vessel Density (ELVD) is calculated by measuring the vessel area under each local window, then normalized with zero mean and unit variance[1]. The local windows are obtained sampling r and θ. Different values of r and θ will tolerate or emphasize different problems with the image quality. In Fig. 3.5 for example, the 4 images on the left ($\theta = 8$ and $r = 3$) have 8 windows each on the centre of VFOV where the macula is located. In this fashion, ELVD features can detect a misaligned fundus image. On the other hand, the ELVD in the 4 images on the right ($\theta = 12$ and $r = 1$)

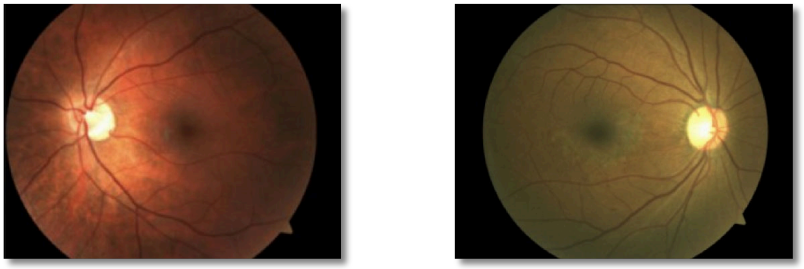


FIGURE 3.6: Pigmentation difference between Caucasian (a) and African American (b) retinas.

will be more robust to macula misalignment, but more sensitive to vessel detection on both vascular arcades.

The idea behind ELVD is to create local windows that are roughly placed in consistent positions throughout different images. In the even or odd columns of Fig. 3.5, note that vessels close to the ON are in the same or nearby local windows, even if images have different FOVs. The power of this new style of windowing is its capability of capturing morphological information about fundus images without directly computing the position of ON, macula or arcade vessels, since these operations are computational expensive and prone to errors if the image has a very poor quality.

3.2.2.2 Luminosity/Colour Information

The analysis of the global colour information of the fundus image can contain useful information for the quality of the image. The method of Lee and Wang (1999) employed the histogram of the grey-level obtained from the RGB image as the only means to describe the image quality. The much more refined method of Niemeijer et al. (2006) uses 5 bins of each channel of the RGB histogram as additional features as input to the classifier. The authors presented results demonstrating that this piece of RGB information improved their classification respect to pure ISI features, even if ISI is representative of most of the retinal structures.

Inspired by Niemeijer et al. (2006), colour information to represent aspects of quality that cannot be entirely measured with ELVD are used, such as over/underexposed images in which the vasculature is visible or outliers with many features that are recognised as vessels.

[1]The zero mean and unit variance is calculated for each feature across all the training images.

All RGB channels are evaluated by computing the histogram for each plane. The histogram is normalised by the size of the mask in order to make this measure scale independent. It is noted that people from different ethnic origin have a different pigmentation on the retina; this aspect is particularly noticeable in the blue and red channel. For example while Caucasians have a fundus with a very strong red component people of African descent have a darker pigmentation with a much stronger blue component (see figure 3.6). In our case this is not an issue because we ensure we have adequate examples of different ethnic groups in our training library.

Also, the HSV colour space is employed as a feature. Only the saturation channel is used which seems to play an important role in the detection of the over/under exposition of the images. The reason is the channel relative independence from pigment and luminosity. Once again, the global histogram is extracted and normalised with the image mask.

3.2.2.3 Other Features

In addition to ELVD and colour information two other sets of features are considered as candidates to represent quality:

- **Vessel Luminosity:** Wang et al. (2001) noted that the grey level value corresponding to the vessels can be used as a good approximation of the background luminosity. They proposed an algorithm that exploits these information to normalise the luminosity of the fundus images. If the vessel luminosity with the same elliptical windows used for the ELVD, we can measure the luminosity spread in the image. This can be particularly useful because poor quality images have often an uneven illumination.

- **Local Binary Patterns (LBP):** Texture descriptors are numerical measures of texture patterns in an image. LBP are capable of describing a texture in a compact manner independently from rotation and luminosity (Ojala and Pietikainen, 1996). The LBP processing creates binary codes depending on the relation between grey levels in a local neighbourhood. In the QA context this type of descriptor can be useful to check if the particular patterns found in a good quality retina are present in the image. This is accomplished by generating an histogram of the LBP structures found.

3.2.3 Classification

The majority of the authors who developed a QA metric for retinal images approached the classification in a similar way (Lee and Wang, 1999; Lalonde et al., 2001; Usher et al., 2003). The training phase consists of creating models of good and poor quality images (in some cases more intermediate models are employed) by calculating the mean of the features of the training sets. For each new retinal image, the feature vector is computed and compared to the two models. Then, it is classified based on the shortest distance[2] to one of the models. This type of approach works reasonably well if the image to be classified is similar enough to one of the models. Also, it simplifies the calculation of a QA metric between 0 and 1 because distances can be easily normalised. However, this approach has a major drawback: the lack of generalisation on images with a large distance from the both models. This problem limits the method applicability in a real world environment.

Niemeijer et al. (2006) are the only authors to our knowledge that approach the QA as a classic pattern classification problem. During the training phase they do not try to build a model or to make any assumption about the distribution of the data. Instead, they label each samples in one of the two classes and train one of the following classifiers: Support Vector Machines (SVM), Quadratic Discriminant Classifier (QDC), Linear Discriminant Classifier (LDC) and *k*-Nearest Neighbour Classifier (KNNC). Finally, they selected the classifier with the best performance (in their case a SVM with radial basis kernel) by testing it with a separate dataset.

Our classification technique is similar to the one of Niemeijer et al., but with two major differences. The first one is that our feature vector is more computationally efficient and it does not need pre-clustering. The second difference is the fact that the classifier needs to output a posterior probability rather than a clear cut classification of a particular class. This probability will allow the correct classification of fair quality images even if the training is performed on two classes only.

3.3 Tests and Results

In this section, a summary of the most significant experiments performed during the development of the ELVD quality estimator are presented. The first section contains an overview of the datasets used. We then show the tests used for an initial evaluation of the QA proposed, the comparison with existing techniques and the choice of the classifier. Then, an analysis on possible optimizations of the feature set is performed. Finally

[2] Distances calculations vary; some use Euclidean distance, others are based on correlation measures.

the final QA system is tested on all the datasets and its computational performance is evaluated.

3.3.1 Material

Few small datasets are employed in the following tests. The list follows:

- **"Image Set 1":** it is composed of 75 images extracted from the Abramoff dataset and labelled as good, fair and poor quality. These images were compiled at the Oak Ridge National Laboratory for the analysis of various aspects of the automatic diagnosis of diabetic retinopathy.
- **"Image Set 2":** this set is composed of 42 images extracted from the Abramoff dataset labelled as good and poor quality. They are good representatives of various aspects of the quality aspects of fundus images. These images were extracted by an expert of the field (Dr. E. Chaum) in order to facilitate the development of the QA system.
- **"Outliers":** it is composed of 24 miscellaneous images containing various types of image outliers, all captured with a fundus camera.

3.3.2 Classifier Selection

In order to select the most appropriate classifier, a series of comparative tests is run on the "Image Set 1" and "Outliers" dataset. The results are compared with our implementation of the QA by Niemeijer et al. (2006). The feature vector used by our classifiers is composed of ELVD with 3 slices and 8 wedges (ELVD 3x8) and the RGB colour histogram with 5 bins per channel.

The testing method used a randomised 2-fold validation, which works as follows. The samples are split in two sets A and B. In the first phase A is used for training and B for testing, then roles are inverted and B is used for training and A for testing. The performance of a classifier are evaluated using the AUC assuming as positive sample a good quality image (TPR/FPR) and vice versa (TNR/FNR).

In the two columns on left, table 3.1 shows the Good/Poor classification results for the "Image Set 1" dataset. All the classifiers using our feature vector have perfect or near-perfect performance in the selection between good and poor class, which is not the case for the Niemeijer et al. method (note that only the good and poor classes are used).

[2]For privacy reasons the ethnicity of the subjects in the Abramoff dataset was not known.

Classifier	Image Set 1		Image Set 1 + Outliers dataset	
	TPR/FPR	TNR/FNR	TPR/FPR	TNR/FNR
Nearest Neighbour	1	1	1	1
KNN (K=5)	1	1	0.99	0.98
SVM (Linear)	1	1	0.92	0.79
SVM (Radial)	1	1	1	1
ISC by Niemeijer et al.	1	0.88	1	0.88

TABLE 3.1: Good/Poor classifier test on "Image Set 1" and "Outliers" dataset. For the first four classifiers the feature vector used is ELVD 3x8 + RGB histogram with 5 bins.

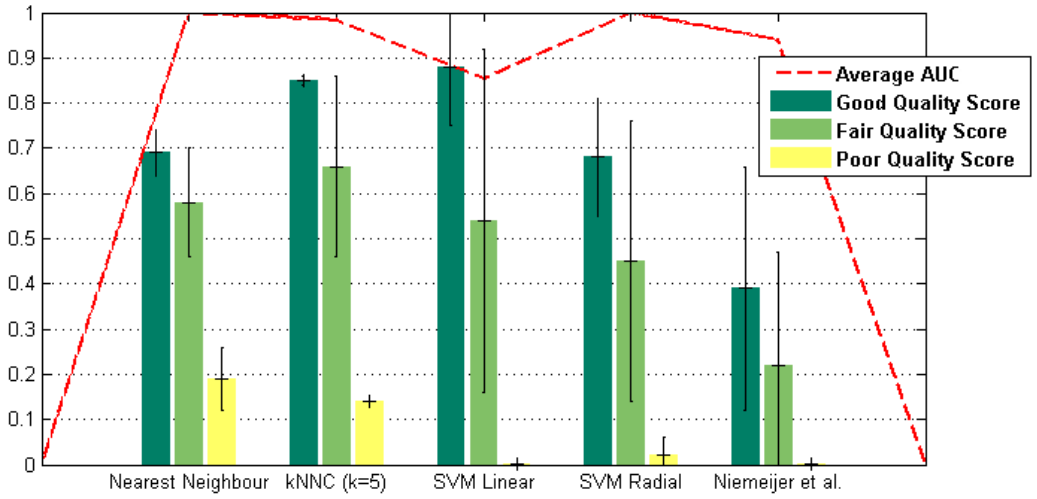


FIGURE 3.7: Classifier scores test on "Image Set 1" dataset. For the first four classifiers the feature vector used is ELVD 3x8 + RGB histogram with 5 bins.

In the two columns on the right, all the Outliers dataset were added as test samples. An outlier image can have an enormous variability, therefore we feel that the training on this type of images might bias the classifier. Ideally, a classifier should be able to classify them as poor even if they are not fundus images as such. In this test, the classifiers performed differently, with the best results are given by Nearest Neighbour classifier and SVM with a radial kernel.

Recall that the aim of this system is to generate a quality score from 0 to 1 to judge the image quality. In order to analyse this aspect, means and standard deviations of the scores obtained are displayed in Fig. 3.7. The classifiers are again trained on Good and Poor class (with 2-fold validation) but the Fair class is added to the testing samples without any explicit training on it, which allows a test of the generalisation of the system. In this test, the classifier with the poorest average AUC (SVM with a linear kernel) is also the one that achieves the best class separation, with an average score separation between Good and Poor classes of more than 0.8. The Fair class in this test has a mean score located at the middle of the scale.

This apparent contradiction makes the selection of the classifier difficult. Therefore another series of tests was run on the more challenging "Image Set 2" dataset. In this case a leave-one-out strategy is used, i.e. the classifier is trained multiple times removing a different sample from the training set and using it as test target each time. This technique allows us to run complete tests using a relative small dataset.

Table 3.2 shows the results obtained employing the same classifiers and feature vector as before. While no classifier obtained ideal performance, the SVM with a linear kernel seems to have a good compromise between AUC and score separation. The small AUC advantages of KNN and Nearest Neighbour do not justify the computational performance issues that these type of classifiers have when many training samples in a high dimensional space are used, and also these classifiers have relatively low score difference between Good and Poor class.

Classifier	AUC TPR/FPR	AUC TNR/FNR	Average Good/Poor score difference
Nearest Neighbour	0.97	0.97	0.51
KNN (K=5)	0.97	0.97	0.51
SVM (Linear)	0.97	0.94	0.76
SVM (Radial)	0.94	0.91	0.54

TABLE 3.2: Good/Poor classifier test on "Image Set 2" dataset. The feature vector used is ELVD 3x8 + RGB histogram with 5 bins (the error bars show the average standard deviation).

3.3.3 Features Selection

The Linear SVM classifier gives good score separation but was not as effective on the identification of outliers. One solution to this problem is the selection of new feature capable of linearising the space. However, the selection of adequate features allowing the SVM hyperplane to split the good quality samples from all the rest is not a straightforward task. Testing all the possible combination of the feature sets mentioned is impractical. Each feature set has many parameters: ELVD 36 (3 sets of radial section and 12 sets of wedges), Vessel Luminosity 36 (same as previously), RGB histogram 80 (all the channel combinations which can be normalised or not and 5 sets of histogram bins), HSV histogram 80 (same as previously) and LBP 4 (2 sets of radii length and 2 sets of LBP codes), for a total of 33 177 600 possible combinations.

Therefore a naive approach to feature selection was adopted. Firstly it is assumed that all feature sets represent independent aspects of the fundus image quality. While this assumption is rather far-fetched, it does allow us to run only 324 tests to check all the possible permutation in each feature set, and also gives a feeling for what features are

worth testing. Table 3.3 shows which are the parameters that achieved the best results for each feature set on the "Image Set 2" dataset. This dataset was chosen since we feel that it is the most authoritative representation of good and poor quality image in most of the different aspects.

Feature Set	Parameters	Avg AUC	Average Good/Poor score difference
ELVD	16 rad. sec. & 6 wedges	0.98	0.74
RGB Hist	4 bins per ch. & mask norm.	0.81	0.51
HSV Hist	5 bins of Sat. & mask norm.	0.85	0.59
Vessel Luminosity	16 rad. sec. & 6 wedges	0.98	0.74
LBP	8 px radius & 8 codes	0.85	0.59

TABLE 3.3: Best results of each independent feature set on the "Image Set 2" dataset. The test is a leave-one-out with a SVM Linear classifier.

If feature sets were actually independent, the ideal feature vector would be composed by all of them with the parameters shown in Table 3.3. However, because there is almost certainly some degree of correlation, various parameters of the feature sets are selected based on their relative AUC and Good/Poor score difference and they are combined together for a total of 16 800 tests.

Optimal results (Avg AUC of 1) and excellent good/poor score separability (0.91) are obtained on "Image Set 2" with a relatively simple feature vector composed of:

- ELVD with 6 wedges and a single radial section
- The mask normalised histogram of the saturation with 2 bins

As it was suspected, the parameters that lead to the best results in this test are not the combination of the parameters found in each independent feature set test (table 3.3). However, they reduce the parameters search space and obtain excellent results with a relative simple combination.

3.3.4 Computational Performance

The performance of the C++ implementation of the ELVD QA is evaluated with a standard benchmarking technique. The complete ELVD QA system is run on 25 images randomly chosen from the "Image Set 2" dataset, during each iteration the time required to run the total system and each separate algorithm is recorded and averaged. All the images are scaled to the common resolution of 756x576 in order to have fairly

Stage	Time (in milliseconds)
Mask Detection	116
VFOV	16
Vessel Segmentation	1920
ELVD	15
Saturation Histogram	25
Classification + Memory Allocation	38
	Total 2130

TABLE 3.4: Relative performance of the different components in the ELVD QA C++ implementation.

consistent measurements. All the test were run on a 3.4 GHz Intel Pentium 4 machine with 2 GB of RAM.

The total time required to obtain a quality score for a single image is 2130 milliseconds. Table 3.4 shows how each system component contributes to the global computational time. The vessels segmentation is by far the main contributor having more than 10 times the computational cost of all the other algorithms summed together. The mask detection and the classification, two possibly expensive operations, are actually quite efficient considering the needs of this system.

For comparison, a global benchmark was run on our implementation of the Niemeijer et al. QA classification (Niemeijer et al., 2006). The result obtained is well over 30 seconds, a time one order of magnitude greater than our approach. This is due to the many filterbanks that must be executed to calculate the raw features and the nearest neighbour operations to obtain the "words". However, the comparison between the two techniques should be taken with a bit of perspective because of the different implementation platforms. In fact the Niemeijer et al. algorithm is implemented in Matlab, a slower language than C++ because of its interpreted language nature. Nevertheless, we should point out that in our tests Matlab uses fast native code thanks to the Intel IPP libraries (Intel, 2007) for all the filtering operations, and these are very computationally efficient regardless of programming language choice.

3.4 Discussion

The new approach described in this chapter was partially inspired by the following techniques: colour was used as features as in the "Histogram Based" technique, the

vessels were segmented as a preprocessing step like in the "Retina Morphology" techniques and the QA was computed by a classifier similar to the one used in the "Bag-of-Words" techniques. Nevertheless, new features were developed and used such as ELVD, VFOV and the use of the HSV colour space. This made possible the creation of a method capable of classifying the quality of an image with a score from 0 to 1 in a period of time much shorter than "Retina Morphology" and "Bag-of-Words" techniques.

Features, classifier types and other parameters were selected based on the results of empirical tests. Four different types of datasets were used. Although none are very large (none contained more than 100 images) they were fairly good representative of the variation of fundus images in terms of quality, camera used and patient's ethnicity. In the literature, the method which seemed to perform best and which had the best generalisation was the one of Niemeijer et al. It was implemented and compared to our algorithm. Our results are in favour of the method presented in this chapter in terms of classification performance and speed. However, while our method has a clear advantage in terms of speed (it runs one order of magnitude faster because of the lower computational complexity), the comparison in terms of classification should be taken with care. In fact, Niemeijer et al. employed a dataset larger than ours to train the system.

However, there were some indirect confirmation of the good performance of the algorithm (which is currently used as help to provide healthcare). First, a strong correlation between a ON localization algorithm performance and QA score was shown by Karnowski et al. (2009). Also, we noted a measurable and increasing trend in the image quality in the clinics that are acquiring fundus images for the telemedicine network described by Li et al. (2011) as shown in Table 3.5.

Month	Images with QA > 0.4	n° of images
January	39.53%	86
February	89.80%	49
March	85.29%	34
April	82.65%	219
May	98.39%	124
June	98.65%	148
July	98.65%	74

TABLE 3.5: Quality assessment trend in a remote clinic

Every time operators were acquiring an image, they were shown the QA score. If the score was below 0.4 they were asked to re-acquire the image, if the second image had again a QA below 0.4 they could chose if submit the image anyway or acquire yet another image. This process seems to have progressively trained the photographers in capturing good quality images. Unfortunately, we did not have a "control group" to

evaluate the relative gain due to the use of QA as feedback to photographers. However, we can assume that in a clinical settings the skills of photographers will increase in a way or in another and the QA qualitatively correlates with this assumption.

Chapter 4

Exudate Segmentation

"Amor y deseo son dos cosas diferentes;
que no todo lo que se ama se desea, ni todo lo que se desea se ama."
"¡Amor, ayuda al deseo, puesto que me pusiste en el!"
- El ingenioso hidalgo Don Quijote de la Mancha, Miguel de Cervantes

Diabetic macular edema (DME) is a common vision threatening complication of diabetic retinopathy which can be assessed by detecting exudates (a type of bright lesion) in fundus images. In this chapter, two new methods for the detection of exudates are presented which do not use a supervised learning step; therefore, they do not require labelled lesion training sets which might be time consuming to create, difficult to obtain and prone to human error. The HEI-MED dataset is introduced which will be used for various experiments throughout the rest of this thesis. We evaluate our algorithm with this dataset and compare our results with two recent exudate segmentation algorithms from the literature. In the tests performed, our algorithms perform better or comparable with an order of magnitude reduction in computational time. Part of this work has been published in Giancardo et al. (2011d,c).

4.1 Introduction

In this chapter, two variations of a new exudate segmentation method that falls into the category of *thresholding methods* which do not require supervised learning are presented (see Section 2.3.2). By avoiding a supervised learning step, we prevent common issues with human segmentation inconsistencies. In addition, many supervised learning methods require large amounts of data which can be very difficult to obtain in this application domain. We introduce a new method to normalize the fundus image and directly compare our method with an implementation of a morphology based technique Sopharak et al. (2008) and another thresholding based technique Sanchez et al. (2009a). In Section 4.2 we introduce the new public dataset used for testing the algorithm; Section 4.3 illustrates two methods published by two other groups which have been implemented and tested on the HEI-MED dataset; Section 4.4 presents the details of two variations of two automatic exudate segmentation methods; Section 4.5 presents the results by comparing them to other two published techniques; finally, Section 4.6 concludes the chapter with a discussion.

4.2 Materials: HEI-MED dataset

The exudates segmentation algorithms have been designed and tested with images from the telemedicine network introduced in Section 2.4.2 and described in details by Li et al. (2011). In January 2010, the central server stored 1907 images of which 75% were healthy and the remaining 25% had some form of retinal condition which required referral to an ophthalmologist. The images were acquired from 910 patients in 971 sessions. This wealth of images are of particular interest for research purposes because it provides an appropriate testbed for lesion segmentation and diagnosis stratification algorithms required for a fully automated eye screening system. We believe that the creation of a dataset based on this pool of data, obtained under true clinical conditions with a broad-based screening protocol, has several advantages over existing datasets as described below.

- **Diverse Ethnicity**: The clinics employing the telemedicine network are mainly located in the Mid-South of the USA, where the ethnic groups are heterogeneous and where retinal pigmentation covers the spectrum generally found in diverse populations. It is important to emphasize this aspect because the appearance of the retinal fundus varies greatly depending on the pigmentation of the retinal pigment epithelium, which is correlated to the ethnic group and eye colour. Depending on this pigmentation, lesions or other type of structures are more or less

FIGURE 4.1: Examples of fundus images in the HEI-MED dataset. (a) African American patients showing clear exudates; (b) Hispanic patient without signs of DME; (c) African American patient that shows choroidal vessels under the pigment epithelium layer and some small exudates; (d) Caucasian patient without signs of DME.

apparent which makes the development of lesion segmentation and diagnosis algorithms that work on a broad spectrum of patients more challenging as shown in Fig. 4.1.

- **Image Consistency**: All the images are captured with the Zeiss Visucam PRO fundus camera, at a resolution of 2196×1958 pixels and with a 45° Field of View (FOV)[1].

- **Quality Assessment**: The image capturing process is vetted by the automatic quality assessment algorithm described in Chapter 3. Every time an image is captured the algorithm is run and the result is shown to camera operators, giving them the chance of taking a new image if required. Thus each image in the dataset has an ELVD quality value stored for comparison.

- **Metadata**: In all the images, additional information is supplied. In addition to the ethnicity and quality metric, age of the patient at image capture, patient gender, type of diabetes, the machine segmented vasculature (employing the method of Zana and Klein (2001)) and a manually generated location of the Optic Nerve (ON).

From the pool of images collected on the server, 169 images representative of various degree of DR/DME were randomly selected. We have verified that all images are of sufficient quality, no patient is duplicated, and a reasonable mixture of ethnicity and disease stratification is represented. Table 4.1 shows the distributions of the ethnicity, DME diagnosis, ELVD quality, diabetes type and patients' age in the dataset. This dataset is publicly available to the research community on the website: http://vibot.u-bourgogne.fr/luca/heimed.php.

[1]Some clinics in the network use a 30° FOV. However, the images used to create the HEI-MED dataset do not contain any image of this type.

Although our algorithm does not use machine learning methods to classify lesions into true positives and false positives, each image of the dataset was manually segmented by E. Chaum, a practising retina specialist. This allowed the evaluation of all the algorithms against the labelled lesions. He identified all the exudation areas and other bright lesions such as cotton wool spots, drusen or clearly visible fluid occurring in fundus images. The manual segmentation and grading of each image was performed on a 17 inch tablet device that can be operated with a stylus, in order to increase the labelling precision and throughput. The software used to perform the task was internally developed with the look and feel of ordinary bitmap painting software. It allows segmentation of the different lesions, revision of already processed images by viewing or amending the existing segmentation, and zooming views. The zooming capability can be somewhat problematic, for while it allows the user to define the contour of the lesion very precisely, it might greatly burden the grader when many very small lesions are clustered together. Therefore, it was agreed to segment clusters of small lesions as a single one if the lesions are not discernible at the first zoom level (a complete view of the image at a resolution of 1280×1024 pixels).

Ethnicity	
African American	104 (62%)
Caucasian	42 (25%)
Hispanic	19 (11%)
Unknown	4 (2%)
DME Diagnosis	
Negative	115 (68%)
Positive	54 (32%)
ELVD Quality Metric	
Poor (ELVD < 0.5)	14 (8%)
Good ($0.5 \leq$ ELVD < 0.8)	31 (18%)
Excellent (ELVD ≥ 0.8)	124 (74%)
Diabetes Type	
Type I	160 (95%)
Type II	6 (3.5%)
Unknown	3 (1.5%)
Patients' Age	
age < 26	5 (3%)
$26 \leq$ age < 43	20 (12%)
$43 \leq$ age < 61	105 (62%)
age ≥ 61	39 (23%)

TABLE 4.1: HEI-MED Dataset Distribution

Using the ground truth labels established by a single ophthalmologist is suboptimal, but in practice was required. The GT coming from different experts permits to average out the human error and to generate a stronger dataset. However, the process of manual lesion segmentation is particularly time consuming and, in our experience, quite difficult to be obtained by busy ophthalmologists. Therefore, we settled for the lesion

labelling coming from a single expert. In any case, the exudates segmentation algorithms developed and tested, do not require a training phase on the lesions, so they are not susceptible to errors and outliers that might be present in the lesion GT. During the testing phase, all the algorithms are tested on the same lesion GT, therefore all the possible errors should not favour any particular technique.

4.3 Exudate Detection Methods for Comparison

In this section the two methods (Sopharak et al., 2008; Sanchez et al., 2009a), which have been implemented and tested on the HEI-MED dataset, are presented. These methods have been selected because they are the two most recent "rule-based" exudate segmentation algorithms presented in peer reviewed journal papers. For rule-based, we intend algorithm that are not explicitly trained on a dataset with a pattern recognition technique, but that requires a limited number of parameters independent of the dataset.

4.3.1 Sopharak et al.

4.3.1.1 Preprocessing

The original image is converted to hue, saturation and intensity (HSI) colour components. The intensity component I is the only one employed for further processing. After removing salt and pepper noise with a small median filter, the contrast of the image is enhanced with an operation defined as contrast-limited adaptive histogram equalisation (CLAHE). Essentially, CLAHE involves a local histogram equalisation on sliding windows that are then combined using bilinear interpolation.

Both ON and vasculature are removed using a series of greyscale morphological operations (Vincent, 1993). In particular, high contrast vessels are eliminated with a closing operation which employs a circular structuring element with a fixed size.

4.3.1.2 Exudate Detection

The sudden variation of the image intensity is captured by calculating the local standard deviation on a sliding window whereby obtaining I_2. I_2 is converted to the binary image I_3 using the Otsu algorithm (Otsu, 1979). In order to cluster together neighbouring pixels which might belong to the same lesion, I_3 is dilated with a binary morphological operation with a circular structuring element. The resulting image is then

flood-filled obtaining I_4. The ON area is removed from I_4 and all the resulting positive pixels Px_{les} appear as an approximation of the lesion candidates.

A soft estimation of the lesion is obtained through morphological reconstruction (Vincent, 1993). The reconstruction is performed using I_4 as a mask image and I as marker image (the locations of Px_{les} are blanked in I). The result of the reconstruction is subtracted from I obtaining the final segmentation result I_{res}. The thresholds used to evaluate I_{res} are $th_{fin} \in \{0 : 0.05 : 1\}$.

4.3.2 Sanchez et al.

4.3.2.1 Preprocessing

Sanchez et al. start the preprocessing of fundus images by extracting the green channel of RGB colour space; in their tests this plane shows the best separability between exudates and non-exudates by the means of the analysis of within-class and between class scatter matrices. In order to reduce the intra and inter image variability, they followed the method proposed by Foracchia et al. (2005). This method identifies the pixels belonging to the background by evaluating the mean and standard deviations on a local window; once the pixels belonging to the background are identified, they are bilinearly interpolated to generate the complete background, which is then used to obtain the normalized image I_{enh}. Fig. 4.2 shows some steps of the process.

The authors notice that the histogram of I_{enh} has a fairly constant unimodal shape across different images where the tail after the peak corresponds to vessel reflections, optic disk and bright lesions. They model the histogram as a mixture of three Gaussians representing respectively: background elements, foreground elements (such as vessels, optic disk and lesion) and outliers. The expectation maximization (EM) algorithm (Duda et al., 2001) is used to iteratively estimate these components.

At the final step of preprocessing, the exudate candidates are extracted by the means of a global threshold that is dynamically selected. The two strongest components are combined and the greylevel corresponding to the deepest concavity on right side of the distribution is used as global threshold. Also, the ON area is detected and removed from the exudate candidates, hence I_{th-ON} is obtained.

4.3.2.2 Exudate detection

The authors argue that edges of exudates are sharper than other bright lesions, such as drusen, regions near the papillary area and other bright artefacts. Hence, they use the

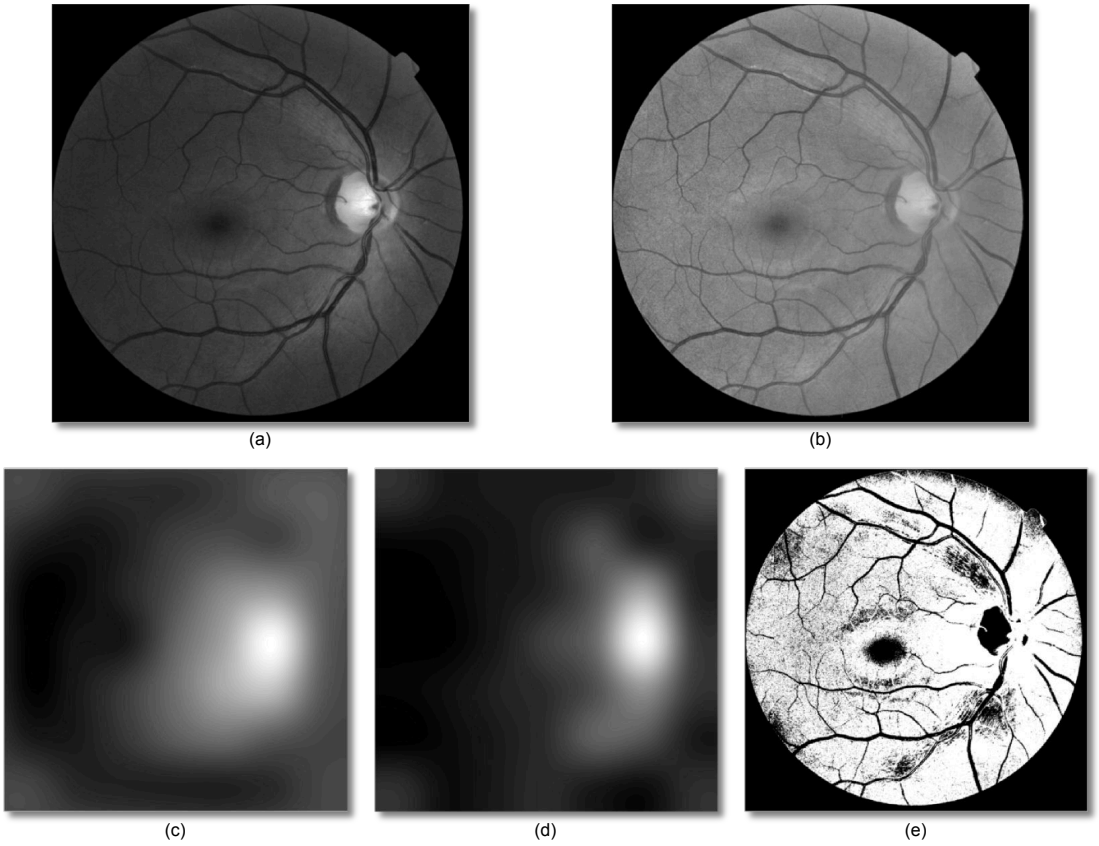


FIGURE 4.2: Background removal by Foracchia et al. (2005). (a) Original Image; (b) Normalized image (I_{enh}); (c) Local mean; (d) Local standard deviation; (e) Point used to estimate the background.

original green plane of the image and apply the Kirsch's edges (Kirsch, 1971) to identify edges at 8 different directions.

Morphological reconstruction and subtraction operations are used to maintain the edges only in the areas corresponding to the candidates of I_{th-ON}. The probability of being an exudate is estimated on a lesion by lesion basis. The binary image I_{th-ON} is clustered with connected component analysis, and for each blob identified the mean intensity is calculated for the corresponding area of the edge image. The higher the value, the higher the likelihood of being an exudate. The thresholds used to evaluate the final output are $th_{fin} \in \{0 : 0.05 : 1\}$.

4.4 Method

The two new methods we present here are identified as Kirsch-based Exudate Detection (KED) and Stationary Waveled Exudate Detection (SWED). They are each presented in

two parts: *Preprocessing* (which is shared by both techniques) and *Exudate detection*. Fig. 4.5 shows an example of a fundus image, the intermediate and final steps of our techniques.

4.4.1 Preprocessing

Our approach uses the green channel I_g of the image and the I_i channel from the HSI colour space. We start the analysis by estimating the background with a large median filter, whose size is $\frac{1}{30}$ the height of the fundus image on I_i. This approach has been used previously (Niemeijer et al., 2005; Fleming et al., 2006b) and has considerable computation performance advantages over other methods (Foracchia et al., 2005). In other median filtering normalization techniques, the background is subtracted from the original image in order to obtain a normalized version. In our approach, we enhance the normalization with morphological reconstruction (Vincent, 1993) as opposed to the more common practise of subtracting the median filtered result, which improves the removal of nerve fibre layer and other structures at the edges of the optic nerve (ON). This is because the shape of the estimated background is more adapted to the original image. The pseudocode of Algorithm 1 illustrates this step. Additionally, Fig. 4.3 compares the background estimated with and without the morphological reconstruction step.

Algorithm 1 Background Estimation

```
1:  function MorphBgEst(I_g)
2:      bgEst ← medianFilter(I_g)
3:      initialise(bgMask)                          ▷ set all the pixels to 0
4:      for y = 0 to height(I_g)−1 do
5:          for x = 0 to width(I_g)−1 do
6:              if I_g(x,y) < bgEst(x,y) then
7:                  bgMask(x,y) ← bgEst(x,y)
8:              else
9:                  bgMask(x,y) ← I_g(x,y)
10:             end if
11:         end for
12:     end for
13:     bgEst_2 ← morphReconstr(bgEst, bgMask)
14: return bgEst_2
15: end function
```

Once the background $bgEst_2$ is estimated, it is subtracted from the original image with signed precision. In the resulting image, the highest peak of the histogram is always centred on zero regardless of the ethnicity of the patient or disease stratification. The histogram shows a clear distinction between *dark structures* and *bright structures* as shown in Fig. 4.4. The former represents the vasculature, macula, dark lesions and other structures and their distribution varies depending on the ethnicity or pigmentation of the patient. *Bright structures* are found in the positive side of the histogram and

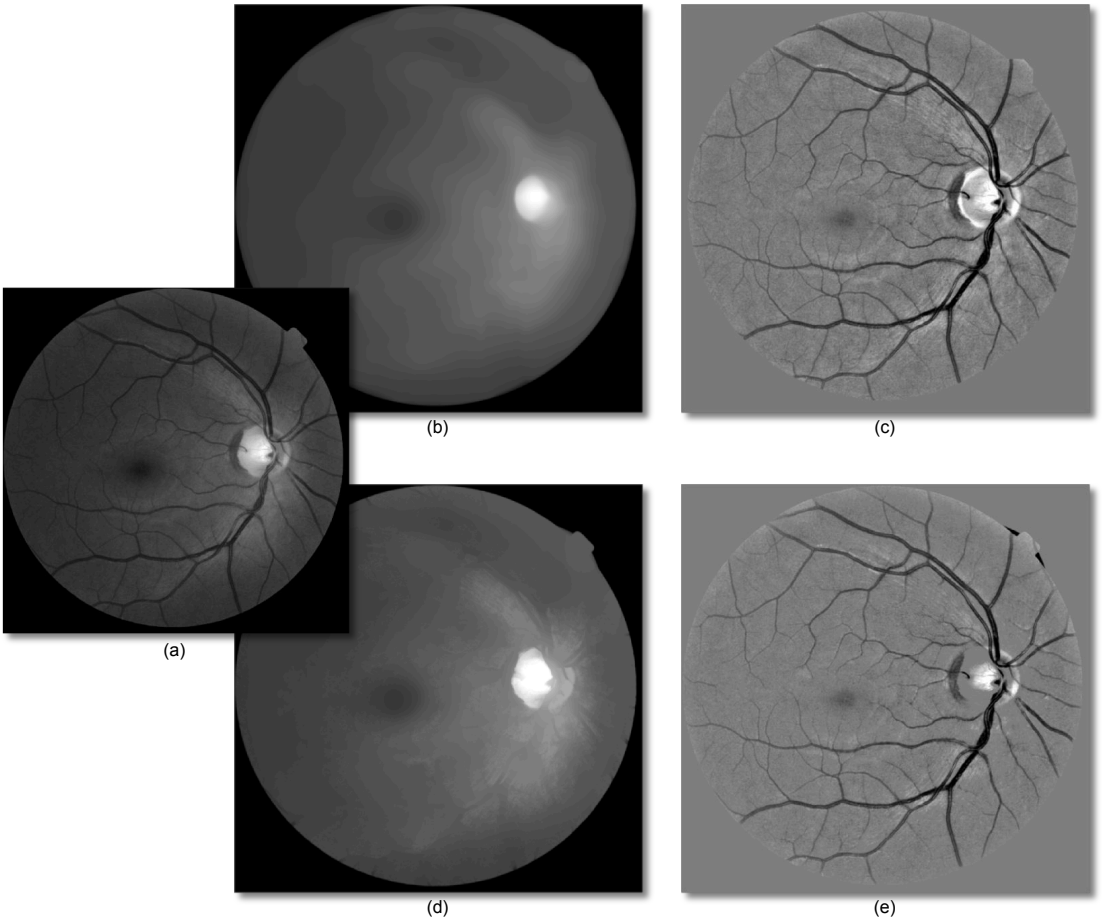


FIGURE 4.3: Background removal based on median filter. (a) original Image; (b) background estimated with median filtering only; (c) image normalized with (b); (d) background estimated with median filtering and morphological reconstruction; (e) image normalized with (d).

include the ON, bright lesions, and other structures. The distribution is fairly constant across different ethnicities.

Because of the characteristics of the normalized image, we can select all the exudate candidates I_{cand} with a hard threshold th_{cand}. This has obvious computational advantages in comparison with model fitting operations which are also more sensitive to suboptimal background subtraction (Sanchez et al., 2009a). In our case we use $th_{cand} = 3$ simply by empirically choosing a value slightly above 0 in order to accommodate small background estimation errors. Fig. 4.5(d) shows an example of the candidates selected. Note that although all the lesions are identified, there are other lesion candidates corresponding to false positives such as nerve fibre layer reflections and other noisy background structures. In instances with large exudates or large clusters of small exudates, the background estimated might contain part of the lesions. This seems to have only the effect of underestimating exudates edges and not missing exudates entirely. We

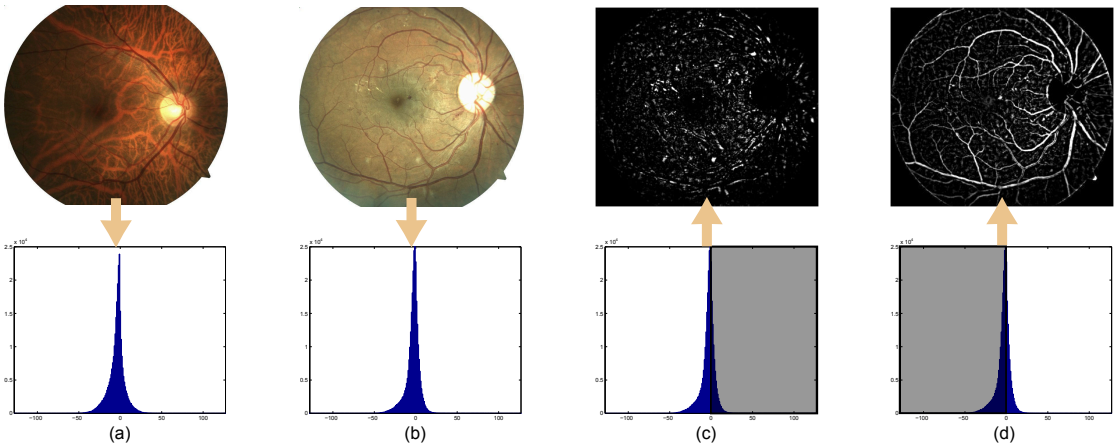


FIGURE 4.4: (a,b) The first row contains original images showing different types of pigments and lesions; the second row shows the image histograms after the normalization process; (c) The bright structures located on the positive side of the histogram; (d) The dark structures located on the negative side of the histogram.

employed the ON location as means to remove the ON from I_{cand} by masking out a region slightly larger than the average ON size. The size of this region was conservatively estimated to be 0.125 times the width and the height of effective retina area. The area outside the viewing aperture (FOV mask) is identified with our fast method (∼16 milliseconds per image) described in 3.2.1.1, which permits to identify the effective area occupied by the image, i.e. the circular area not masked out by the black FOV.

In our algorithms (and in the ones implemented for comparison), we made use of the ON location present in the HEI-MED GT. In this way it is possible to compare the effectiveness of the lesion algorithms by themselves and not the performance of the different ON detectors used. In Section 6.3, an automatic ON detection technique will be included.

4.4.2 Exudate Detection

The exudate detection is performed by assigning a score for each exudate candidate. The exudate candidates are selected by running a 8-neighbour connected component analysis on I_{cand}. We have implemented two ways to assign this score, one based on *Kirsch's Edges* (Kirsch, 1971) and the other based on *Stationary Wavelets*. Both methods seek to take advantage of the higher inner and outer edge values of exudates in comparison to non-exudate structures.

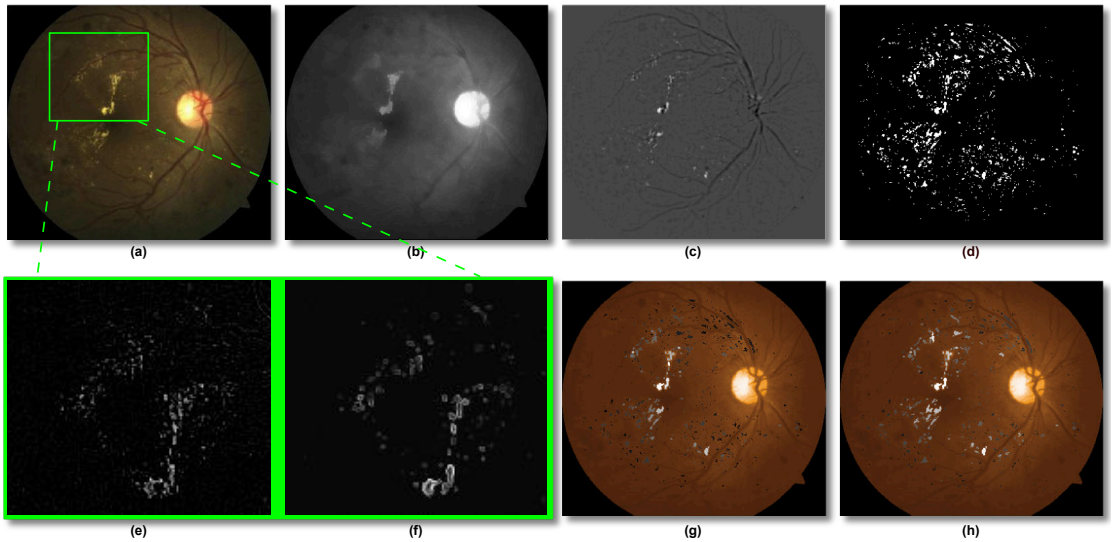


FIGURE 4.5: (a) Example of a fundus image of the dataset used. The square represents the area shown in (e) and (f); (b) Estimated background, $bgEst_2$; (c) Image after background subtraction; (d) Initial exudates candidates, I_{cand}; (e) Image detail of Stationary Haar Wavelet Analysis; (f) Image detail of Kirsch's Edges Image Analysis; (g) Exudates Probability with Wavelet (I_{wav}) overlaid on a enhanced version of the original image (to improve the contrast with the probabilities). White corresponds to a probability of 1.0 of being a true exudates, black to a probability of 0; (h) Same as (g) but showing the Exudates Probability with Kirsch's Edges, I_{kirsch}.

4.4.2.1 Kirsch's Edges (KED)

Kirsch's edges emphasis the external edges of the lesion candidate. This edge detector is based on the kernel k (shown below) evaluated at 8 different directions on I_g. The kernel outputs are combined together by selecting the maximum value found on each pixel output. The result is stored in the final I_{kirsch} image.

$$k = \begin{bmatrix} \frac{5}{15} & -\frac{3}{15} & -\frac{3}{15} \\ \frac{5}{15} & 0 & -\frac{3}{15} \\ \frac{5}{15} & -\frac{3}{15} & -\frac{3}{15} \end{bmatrix} \tag{4.1}$$

The average edge outputs of I_{kirsch} under each lesion cluster are calculated and assigned to the lesion in its entirety. The thresholds used to evaluate the final output are $th_{fin} \in \{0 : 0.5 : 30\}$.

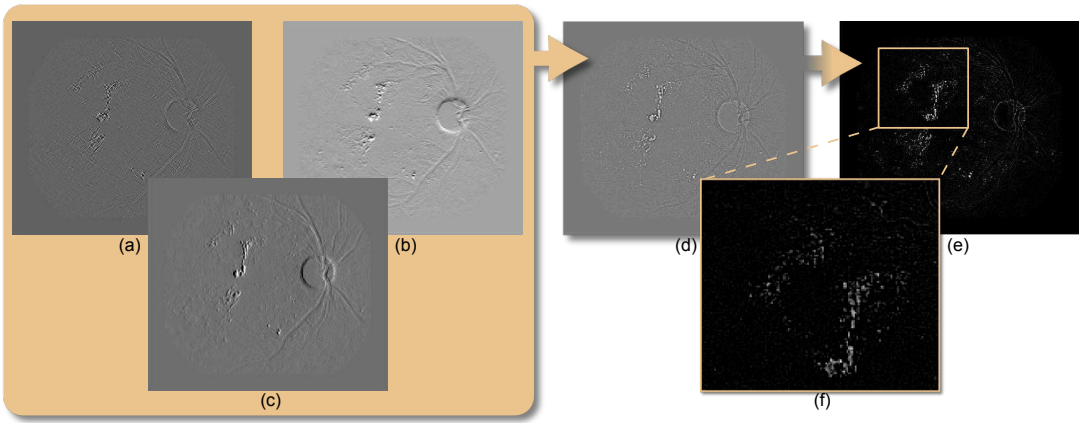


FIGURE 4.6: Haar stationary wavelet analysis of the image shown in Fig. 4.5(a). (a,b,c) Respectively diagonal, vertical and horizontal coefficients of the 2° level decomposition; (d) Image reconstructed using uniquely the coefficients of (a,b,c); (e) Exudate enhanced image by histogram thresholding; (f) Details of (e) highlighting an area showing exudation.

4.4.2.2 Stationary Wavelets (SWED)

Wavelet analysis is a powerful multi-resolution signal analysis technique that has many applications, such as denoising or compression. In comparison to a traditional Fourier analysis, a wavelet approach has many advantages, the most prominent being the fact that wavelet functions are localized in space and that their scale can vary. In order to analyse an image (or any other signal), a mother wavelet must be chosen. A mother wavelet is essentially a basis function which obeys the different properties of wavelets and ideally emphasizes the relevant aspects of the signal of interest. From this mother wavelet a series of scaling and wavelet functions are derived which are able to decompose the image at different scales. In 1-D, two approximation signals are generated for each scale level: one containing the details and the other the basis. In 2-D, apart from the basis, three different details are generated for each scale level: vertical, diagonal and horizontal. For more information, Mallat (1999) provides an introduction on the topic of wavelet analysis.

Quellec et al. (2008) proposed a method for the detection of retina microaneurysms in the wavelet space. We try to capture the strong peak at the centre of exudates by developing a method that employs a similar wavelet analysis, but that evaluates the results in image space. A stationary Haar wavelet analysis is performed up to the second level on I_g. The process is inverted maintaining the last vertical, diagonal and horizontal details only (see Fig. 4.6(a-c)), as these are the wavelet coefficients that seem to contain most of the foreground structures. By transforming back to the image space we obtain a background-less image I_{wav} (see Fig. 4.6(d)). It is interesting to note that

the distribution of I_{wav} has similar properties as the image obtained during the image normalization phase (i.e. centred at 0 and with the exudates located on the positive side of the histogram). Hence, we set to 0 all the reconstructed pixels that correspond to the negative side of the histogram. As it can be seen in Fig. 4.6(e,f), we are able to enhance the response of exudates particularly at their central areas. This can be better appreciated by comparing Fig. 4.6(f) with Fig. 4.5(d) which are magnification of the same image area. Similarly to the previous approach, we evaluate the candidate by detecting variation in the I_{wav} area which corresponds to each lesion cluster of I_{cand} as follows.

$$wavScore = \frac{max(px_{wav}) - min(px_{wav})}{max(px_{wav})} \tag{4.2}$$

where px_{wav} are the pixels of I_{wav} corresponding to a lesion candidate cluster. The thresholds used to evaluate the final output are $th_{fin} \in \{0 : 0.05 : 1\}$.

4.5 Results

In our results, we compare the KED and SWED technique with our implementation of Sopharak et al. (2008) and Sanchez et al. (2009a). In all instances, we evaluate the performance of the lesion segmentation algorithms on a lesion by lesion basis for each image. The analysis starts from the labelled image in the dataset and each lesion is compared to the automatic segmentation one by one. A lesion is considered a true positive (TP) if it overlaps at least in part with the ground-truth; a false negative (FN) if no corresponding lesion is found in the automatic segmentation; a false positive (FP) if an exudate is found in the automatic segmentation but no corresponding lesion has been manually segmented. In the evaluation of the segmentation, we did not employ any true negatives (TN). As such, we avoid any evaluation of specificity, which is inherently high in images where the lesions represent a very small percentage of the total image area. In order to compare the methods fairly, all the images in the dataset have been resized to the resolution used in the original papers (but maintaining the original width/height ratio). It should be noted that in the tests of this chapter all the FP corresponding to bright lesions other than exudates have not been taken into account for the evaluation of TPs or FPs. In fact, confounding a bright lesion such as a cotton wool spot with exudates is not really problematic since they are still a retinal abnormality. However, this was done only for the comparative evaluation of these exudate segmentation algorithms, in the computer-aided diagnosis step presented in Chapter 6 the original segmentation was used for a completely unbiased evaluation.

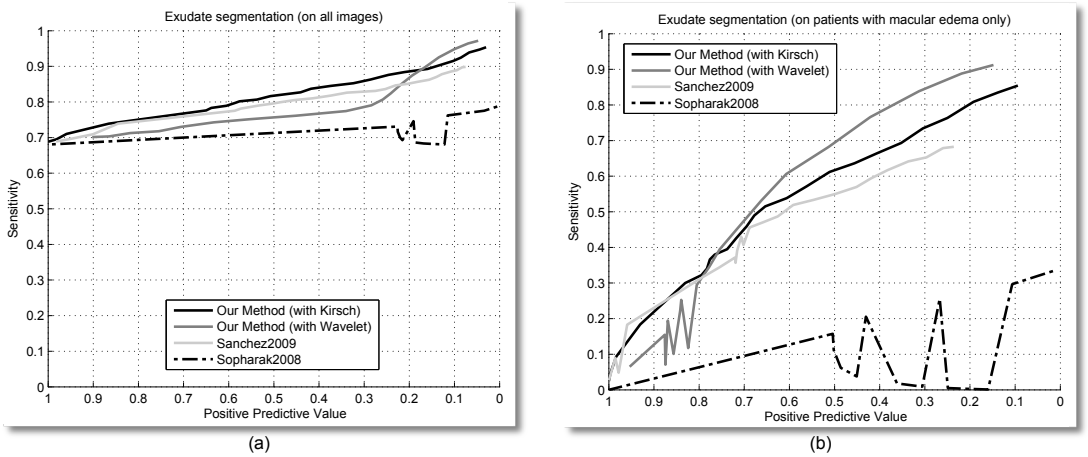


FIGURE 4.7: FROC curves. (a) the exudate segmentation performances on all images; (b) the performance on the images diagnosed with DME only.

Fig. 4.7(a) shows FROC analysis of the implemented algorithms on the entire dataset. It can be noted how our two methods and the technique developed by Sanchez et al. performed comparably, with our technique (with Kirsch's edges) performing somewhat better. Similar results are shown in Fig. 4.7(b). In this case only the 54 images with DME are used in order to have a better evaluation of the segmentation performance. In this case both of our methods seem to have a higher sensitivity than the other two methods implemented.

Since exudation that is close to the fovea it is more likely to cause vision loss, we tested the performance of the Kirsch's edges method in this region (Fig. 4.8(a)). The algorithm performance improves because the macular pigment in the area around fovea tends to be darker than the rest of the image, therefore exudates have more contrast, and there are fewer FPs due to the removed structures near the larger vessels.

We note that the true goal of automated screening is a diagnosis of the patient condition, and lesion segmentation is a step towards this goal. Consequently, the detection of every exudate in an image with many may not be as important as finding a single significant lesion on a image with only one of them. Therefore, we have evaluated the algorithms on their ability to discern patients with or without DME by employing the hard threshold of a single lesion. If one or more exudates are found, the image is diagnosed with DME, otherwise the patient is classified as being negative. Fig. 4.9 shows the results of this experiment by the means of a standard ROC analysis on our dataset. This is possible because we are classifying the image DME condition and not the lesion segmentation, therefore we can employ TN. Again, our methods seem to perform better or comparably to the other algorithms. Although the area under the ROC

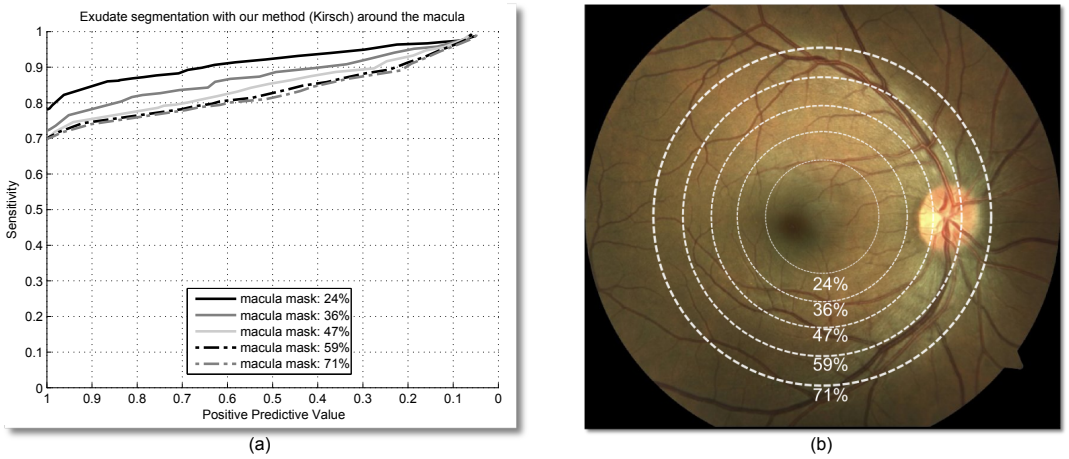


FIGURE 4.8: Performance of our method based on Kirsch's edges around the macula. (a) FROC curves; (b) Areas considered.

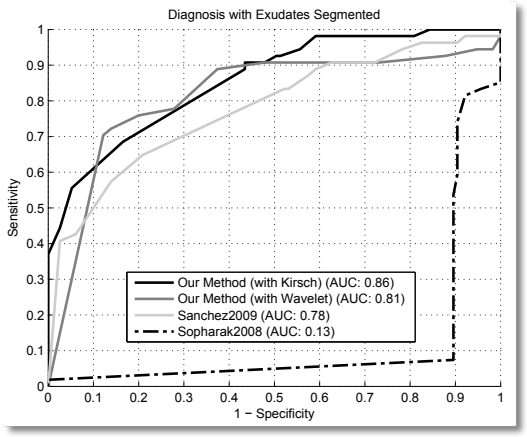


FIGURE 4.9: Naive DME diagnosis based on lesion segmented, i.e. an image is considered positive if it shows at least a lesion.

curve (AUC) is highest for our method with Kirsch's edges, the technique with wavelet shows a higher sensitivity at a higher specificity, a useful aspect for the development of a automatic DME screening tool.

The reason for the suboptimal performance of the algorithm by Sophorak et al. is very likely to be due to the choice of parameters. In our implementation we refrain to change anything from the technique described in their paper, however the many parameters do not seem to be suited to our dataset, which, assuming a correct implementation, would explain the discrepancies between our results and the ones presented in their original paper.

The computational performance are evaluated on a Dual Core 2.6 GHz machine with 4

GB of RAM. All the algorithms are implemented in unoptimised Matlab code. Our methods require 2.4 and 1.9 seconds per image for the Wavelet and Kirsch's edges methods, while the methods by Sanchez et al. (2009a) and Sopharak et al. (2008) require 39 and 36 seconds respectively. A large reason our approaches are more computationally efficient is the high computational cost of the bilinear image interpolation and expectation maximization step in Sanchez et al. (2009a), and the calculation of the standard deviation for each pixel for Sopharak et al. (2008).

4.6 Discussion

We have evaluated the global segmentation performance with the real distribution of patients in a screening setting and only the patients showing signs of exudation. The results are particularly encouraging especially given the comparison with the other techniques by Sopharak et al. and Sanchez et al. The method by Sanchez et al. is somewhat close to our tests, however, the image normalization procedure gives a substantial computational advantage to our method. The median filter with morphological reconstruction approach maintains a good contrast of the foreground structures by limiting the effects of the noise due to nerve fibre layer reflections and other small artefacts. In addition we have evaluated our algorithms ability in identifying patients with DME based on the segmentation of one or more lesion in a fundus image. This is a simplistic method for DME diagnosis, but does provide a baseline of the possible screening performances that can be achieved employing the output of the segmentation as a classification feature.

Many other exudate segmentation methods have been published throughout the years using a variety of datasets and evaluation methods. This makes a direct comparison almost impossible as shown in Table 2.4 and as emphasized by many of the authors themselves. We note this issue in this chapter as we found that our implementation of two algorithms did not perform as well as in the respective datasets employed by the original authors. This issue makes the HEI-MED dataset, which is publicly available for algorithm evaluation purposes, a very relevant tool for future work in this area.

Chapter 5

Microaneurysms Segmentation

"Che io forse abbia amato tanto la sigaretta per poter riversare su di essa la colpa della mia incapacità? Chissà se cessando di fumare io sarei divenuto l'uomo ideale e forte che m'aspettavo? Forse fu tale dubbio che mi legò al mio vizio perché è un modo comodo di vivere quello di credersi grande di una grandezza latente"

- La Coscienza di Zeno, Italo Svevo

The localization of microaneurysms is a key component for the automatic detection of DR. In this chapter, we propose a new microaneurysms segmentation technique based on a novel application of the Radon transform, which is able to identify these lesions without any previous knowledge of the retina morphological features and with minimal image preprocessing. The algorithm has been evaluated on the Retinopathy Online (ROC) Challenge public dataset, and its performance compares with the best current techniques. The performance is particularly good at low false positive ratios, which makes it an ideal candidate for diabetic retinopathy screening systems. Part of this work has been published in Giancardo et al. (2010a, 2011b).

5.1 Introduction

As illustrated in Section 2.1.4.1 microaneurysms (MAs) are a common and often early manifestation of DR. MAs are the primary target lesions for laser treatment of DR. As such, the MA detector is an attractive candidate for an automatic screening system able to detect early findings of DR.

From an image processing standpoint the automatic detection of the microaneurysms presents various challenges. First, MAs colour and size are the same as the vessels (in fact they are small vessels themselves). Microaneurysms have a variable size and often they are so small that can be easily confused with image noise or retina pigmentation variation (or vice versa). In fact, even expert ophthalmologists do not always agree whether a red lesion is a microaneurysm of some type or other structures. The only way to be certain is through fluorescein angiography, an invasive procedure which involves the injection of a contrast agent in the patient. Small dot hemorrhages are types of similar looking structures that, fortunately, are a symptom of DR too, hence for an automatic detector of MAs is not useful to distinguish between the two, especially if the aim of the algorithm is patient screening for referral to ophthalmologists.

In Section 2.3.1, various algorithms for MAs detection were introduced. In particular, the techniques that have participated in the ROC challenge to date were presented. The top performing techniques (Antal et al., 2010; Niemeijer et al., 2005) are based on a combination of multiple MAs localization algorithms that require various intermediate steps such as vessel segmentation and multiple classifiers. In their approaches, the authors come to the conclusion that a combination of multiple MAs localization algorithm showed great improvements if they use different properties of the MAs in their detection.

We present a MAs localization algorithm based a new, novel set of Radon-space features. The algorithm is not based on a combination of different techniques but it reaches performances comparable to the best two algorithms (which are much more complex). The features exploit the round Gaussian-like structures of MAs. The Radon Transform is well known in the Medical Imaging field as a technique to reconstruct a two or three dimensional object from its projections (Resnick, 1985). While its usage for other purposes is uncommon, it is possible to find its application in image compression (Kingston and Autrusseau, 2008) and in texture characterization (Liu et al., 2009). By generating the features in the Radon space, it is possible to exploit its compact representation of Gaussian-like structures of variable size in a sliding window, even when noise or vessels are present. In a previous work (Giancardo et al., 2010a), we derive an operator from these features in a rule-based fashion, however the detection performance was

only average. In this chapter, the samples are classified through Principal Component Analysis (PCA) and a Support Vector Machine (SVM) which increase the performance substantially. This method is streamlined, does not require a vessel segmentation and is easily trainable.

In Section 5.2 we present the details of the algorithm developed; Section 5.3 discusses the characteristics of the ROC dataset and our samples selection approach; Section 5.4 presents the results; finally, Section 5.5 concludes with a discussion of the method and results.

5.2 Methods

The algorithm was developed on 5 images selected from different datasets. One of these images came from the ROC training set. A total of 122 window samples were manually selected with the following distribution: 60 windows containing MAs, 38 windows containing vessels and 24 windows containing only background. This samples are referred as *feasibility study data*.

Our method starts with a very conservative candidates selection phase, in which we rule out the obvious areas that cannot contain any type of dark lesion. This phase is required only to avoid unnecessary computations. Then, we split the image into small windows and we proceed with the creation of the Radon based features. In the final phase, we assign a score representing the likelihood of containing a MA to each window in the image.

5.2.1 Radon Space Characteristics

The algorithm lays its foundations on one observation: every roughly circular Gaussian shape appears as a continuous chain of cliffs if observed in the Radon space. The morphology of the ridges describes three properties of the circular Gaussian shape: its relative location in the window, its size and intensity. The location in the window can be deduced from the shape of the crest, which is a straight line if the Gaussian is located at the centre of the window or another type of curve if the Gaussian is otherwise situated in the window (see Fig. 5.1). The size and intensity are given respectively by the width and height of the cliffs. Fig 5.1 shows how this cliff-like structure is insensitive to a conspicuous amount of noise.

There are other advantages to the analysis in Radon space. When detecting the targeted structure in a window with other techniques based on template matching, the target

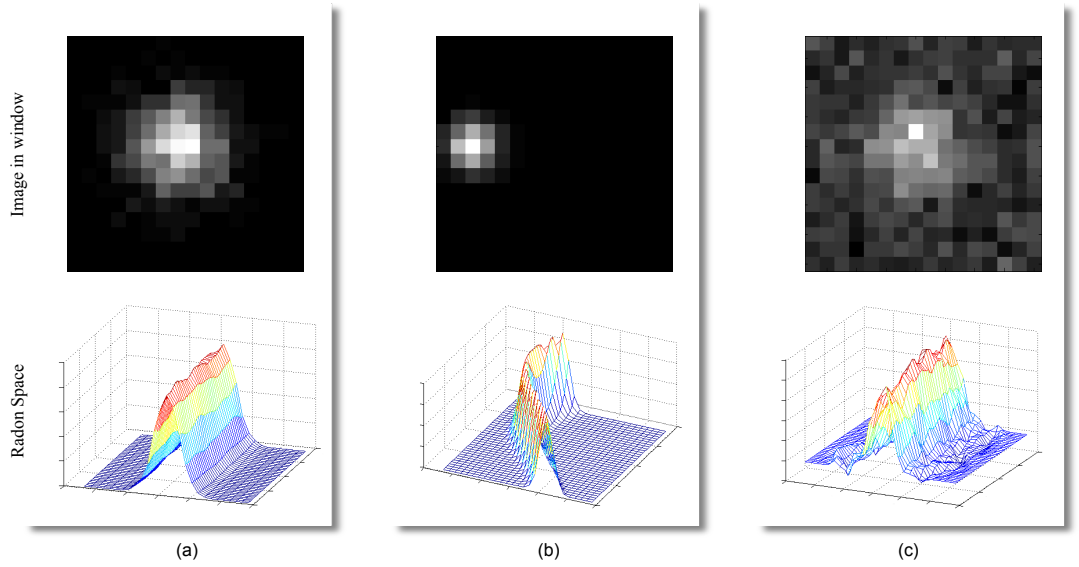


FIGURE 5.1: Illustration of the Radon space appearance for different Gaussian structures. (a) Gaussian intensities centred in the window; (b) Gaussian intensity off-center; (c) Gaussian intensities with considerable noise.

must be the most prominent object in the window. Hence, an algorithm based on this type of technique needs to go through a training phase to find a window containing the target with enough space to contain the support background, but sufficiently small to avoid any external object. Given that MAs are often very close to vessels this might not be possible in all instances. Similar problems occur when a template function is dynamically adapted to the window using least mean square or other regression approaches. On the other hand, Radon-based analysis can identify a Gaussian-like object in a window containing another structure with a higher pixel values. If the Gaussian object is at the center of the window, it will maintain the continuous cliff-like appearance, and the other object will appear as a protrusion which might be connected or disconnected to the central structure. In the context of microaneurysms detection this aspect is of a particular importance since many lesions of this type tend to appear close to large vessels or capillaries.

5.2.2 Candidates Selection

The green channel I_g is extracted from the original RGB image and resized (with bilinear interpolation) such that the original ratio height/width remain unmodified and the new width is 768 pixels. The black background around the FOV (FOV mask) is detected with the method described in Section 3.2.1.1 and stored in I_{mask}. I_g is cropped based on the redundant rows and column of I_{mask} in order to maximize the FOV. Then,

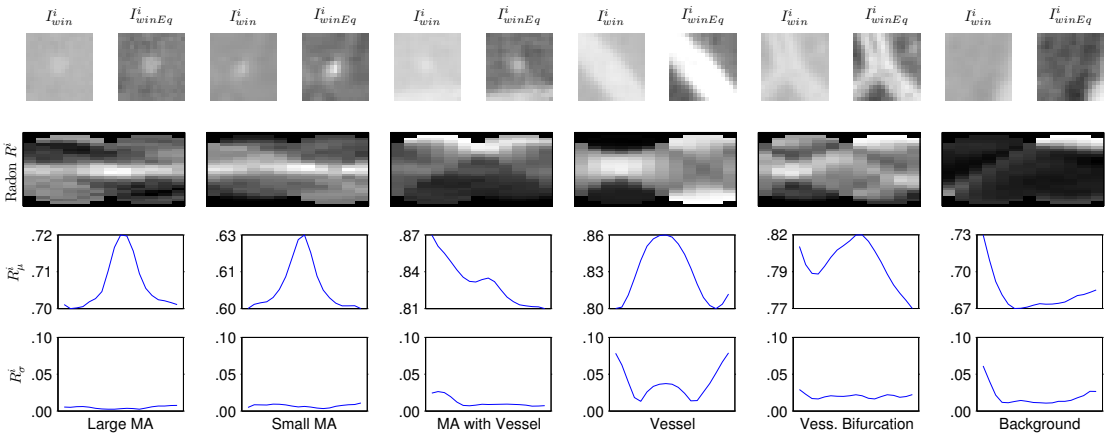


FIGURE 5.2: Visualization of Radon transform analysis. Each column represents the analysis for a single window. In the top row, the original (I^i_{win}) and the equalized (I^i_{winEq}) windows are shown; in the second row, the normalized radon space (R^i) is shown; in the third row, the vectors R^i_μ are displayed (at different scales); in the last row, the vectors R^i_σ are shown.

the background is estimated similarly to the method of Cree et al. (2005). First, an initial background estimation is calculated by the means of a large median filter on I_{gi}, the inverted version of I_g (see Fig. 5.4.b). The dimension of the median filter is 4% the size of I_{gi}. The background image is subtracted from I_{gi} obtaining an image whose distribution is naively assumed to be Gaussian and normalized with $\mu = 0.5$ and $\sigma = 0.2$. All the pixels not laying between 0 and 1 are considered "outliers" and changed to the nearest valid value. While the value of μ was chosen to be in the middle of the allowed range, the value of σ was empirically derived from the 5 images employed in the feasibility study, so that a high percentage of the image pixels (∼95%) would stay in the allowed range. I_{norm} is the image generated by this type of normalization, which allows to compare the pixel values across images with different pigmentation, illumination and contrast. On the top row of Fig. 5.2, details of I_{gi} (I^i_{win}) are shown together with the normalized version (I^i_{winEq}).

The pixels selected as candidates are those having a value greater than a hard threshold $th = 0.58$ and that do not lie on I_{mask}. th is selected to be a very conservative value such that all the areas corresponding to the MAs of the 5 images in the feasibility study are included.

5.2.3 Radon-based Features

The Radon based features are calculated on I_{gi}. This image did not undergo any type of preprocessing apart from cropping, resizing and pixel value inversion operations. By employing I_{gi} directly, we avoid the small artefacts that are inherent to the background subtraction operation, especially on images with a substantial JPG compression such as the ones found in the ROC dataset.

The image is partitioned in 5×5 pixel windows in order to form a grid-like pattern. Each window is deemed as valid if it contains at least one candidate pixel. Each valid window is centered on the pixel with the highest value in its local neighbourhood, in order to have the suspected MA in the middle of the window. The Radon transform is calculated on a 17×17 neighbourhood (I^i_{win}), with scanning angles between 0° and 165° spaced of an interval of 15°. Because of the non-isometric support (the window is a square), at some angles the Radon projections will go through an unequal number of pixels, which leads to an image containing coefficients biased towards certain locations. We overcome this "Radon transform bias" by normalising each projection in each angle for the number of pixels the projection ray goes through; effectively making each coefficient in the Radon space the mean value of the pixels crossed by each projection rays for each angle. The normalization can be obtained efficiently by initially applying the Radon transform on an image containing only ones and having the same size of the sliding window, then using it to scale the coefficients obtained with the Radon transform. Fig. 5.2 shows some examples of the Radon space obtained (R^i).

We attempt to capture the characteristics of the Radon space that separate the MAs from other dark structures, such as vessels or pigmentation noise, with a straightforward analysis by using the equations

$$R^i_\mu(x) = \frac{1}{\phi}\sum_{n=1}^{\phi} R^i(n,x) \quad 0 \le x < \rho \tag{5.1}$$

$$R^i_\sigma(x) = \sqrt{\frac{1}{\phi-1}\sum_{n=1}^{\phi} [R^i(n,x) - R_\mu(x)]^2} \quad 0 \le x < \rho \tag{5.2}$$

where i is the index of a window, R^i is the window in the normalized Radon space having on the horizontal axis the different angles of projections and on the vertical one the number of projections for each angle. ϕ is the number of projection angles and ρ is the number of projection rays. Note that R^i_μ is a vector containing the mean across the rows of R^i. In the case of a MA, it will have a strong maximum in the middle, even if another prominent structure (such as a vessel) is at the periphery of the window I^i_{win}. Since we want to use R^i_μ as features for classification purposes, we need to make these measurements as homogeneous as possible. The fact that the Radon transform is

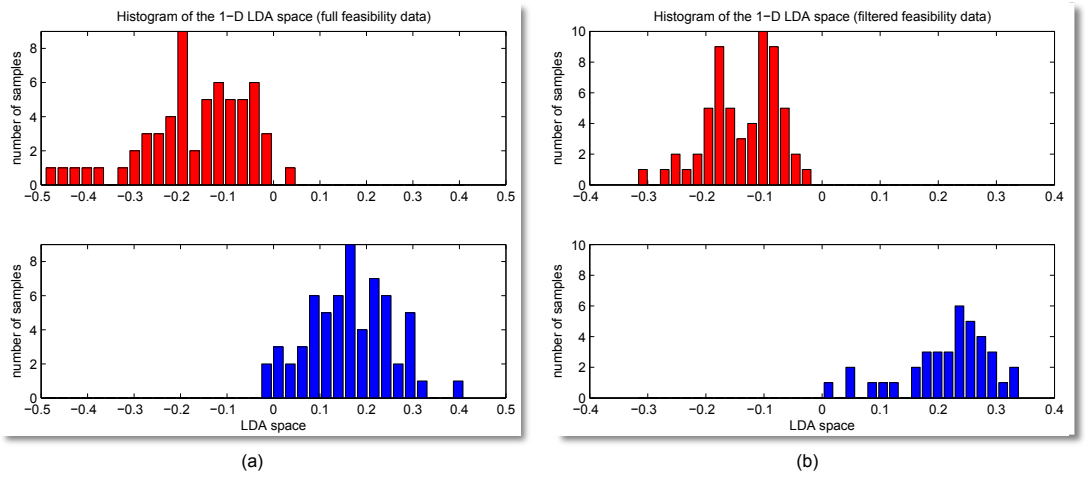


FIGURE 5.3: LDA space histograms of MA samples. Above the samples containing MAs, below the other samples. (a) full feasibility dataset; (b) feasibility dataset filtered with the criteria described in Section 5.2.2.

calculated on a non-normalized image might create problems in this regard. Therefore we employ the first derivative $\frac{d}{dx}R^i_\mu(x)$, which is able to capture the location of maxima and minima without using absolute values (that differ depending on the background pigment, contrast and illumination).

In some instances, there is a strong resemblance between MAs and vessel bifurcations. By employing R^i_σ (i.e. the standard deviation across the rows of R^i) we can determine if the central crest of R^i is constant or not. In the first case, R^i is very likely to contain a MA, but in the second case some type of vessel bifurcation. We add R^i_σ to the feature vector unchanged because the standard deviation is inherently independent of different types of backgrounds. Fig. 5.2 shows the various analysis steps for different sample classes.

Finally, the feature vector is composed as $F^i = \begin{pmatrix} \frac{d}{dx}R^i_\mu \\ R^i_\sigma \end{pmatrix}$.

In order to evaluate the linear separability of the feature vector on the feasibility data, Linear Discriminant Analysis (LDA) is used. Fig. 5.3(a) shows an histogram of the feasibility data samples projected on a 1-dimensional LDA space. The separability is really good especially considering that some of the samples are quite challenging and that no samples have been discarded with the candidate selection phase. Instead, Fig. 5.3(b) shows how the separability becomes ideal when the feasibility data is filtered with the criteria described in the candidate selection phase (Section 5.2.2).

5.2.4 Classification

The original feature vector F^i has 37 dimensions (this number may vary depending on the implementation of the Radon transform). The samples are normalized so that the samples distribution has 0 mean and a standard deviation of 1 across each dimensions. Then, all F^i are projected to an hyperplane of 10 dimensions obtaining F^i_{pca} through Principal Component Analysis (PCA). During the feasibility study, we have estimated that this dimensionality reduction maintains 95% of the original variance, hence little information is lost during the process, making possible the reduction of the data dimensionality. At the feasibility study stage, this was confirmed by a 10-fold classification test: the SVM classifier obtained a score of 0.89 of the Area Under the ROC Curve (AUC) on the original F^i and a score of 0.96 AUC after PCA dimensionality reduction.

As already mentioned, the F^i_{pca} is classified with a SVM. We employed a 3 degrees radial basis kernel with $\epsilon = 0.001$ and the estimation of the probability with the method implemented in the LIBSVM library by J.C. Platt (Chang and Lin, 2001).

The calculation of the probability of being a MA is a combination of the SVM probability output and the average grey level at the centre of the equalized window of I^i_{winEq} 3×3 neighbourhood. The two probabilities are combined by a multiplication as suggested by the unnormalized Bayes rule.

The final step is a non maxima suppression in the matrix containing the MA probabilities, so that neighbouring MAs are not considered separate detections but a single one. This is performed with a morphological closing operation with 12×12 structuring element, followed by a blob analysis. Fig. 5.4 shows an example of the MA detection.

5.3 Materials and Training Strategies

We evaluated the MA detection algorithm on the ROC dataset. In this dataset, four retinal experts annotated all the small red lesions (MA and round hemorrhages) by labelling them as MA or irrelevant lesions. The set was divided in two: training and testing (composed of 50 images each). For the former, the global judgement of the experts was publicly released via an XML file, which did not contain the source of the decision but only the global labels combined together with an OR operation. This maximized the global sensitivity at a price of a certain number of false positives. The lesion labels on the test set were withheld in order to avoid training on the testing set. In this case, the gold standard of the experts' judgement was created with a voting system for a better trade off between sensitivity and false positives.

A novel training strategy was adopted for the classifier. First, we hand selected all the unambiguous MAs from the ROC training set and used them as positive examples. Then, we manually selected a handful of negative examples and trained a "baseline" classifier. At this point, we employed an on-line training technique to update the classifier with negative examples. This is possible through a GUI (that we developed) that is able to add a negative sample, train the classifiers and show the detection in real time. We picked an image showing a few MAs (from the training set), and we added negative samples up until we were satisfied with the detection. At this point the classifier is trained and ready to be evaluated.

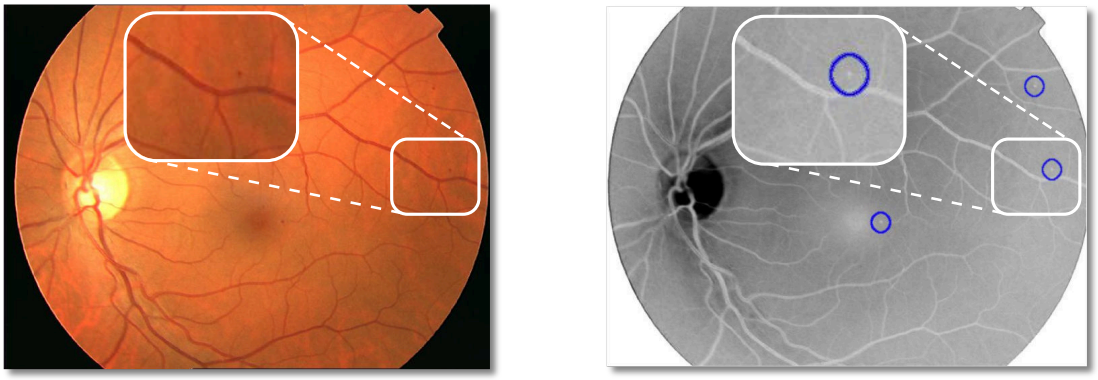
This approach has considerable advantages over a classical training procedure based on manual labelling procedure. It allows us to add only the negative examples that are effectively useful to the MAs classification leaving the uncertain structures out of the training set. By minimizing the samples required, we avoid overtraining and greatly simply the whole training process. We trained the classifier used for these tests in less than 20 minutes by adding 80 negative samples.

In fact, one of the biggest problem for MAs detection is the selection of the samples as positive and (especially) negative examples. MAs can be very small and easily confused with image noise or pigmentation changes. Even when graders have the option of labelling areas as "uncertain", the problem is alleviated but far to be solved. Various authors mentioned this problem also for the ROC dataset, that has been graded by 4 experts and has the option to label lesions as "uncertain" (Zhang et al., 2009).

Other authors are aware of this problem and indicate various solutions. Lazar and Hajdu (2011) employ a pure rule based detection in order to avoiding any direct training on the dataset. Quellec et al. (2008) employs a very computationally expensive genetic algorithm to calculate the ideal parameters of a model based detection on wavelet space. Recently Quellec et al. (2011), re-approached the MAs detection as a classification problem but employing multiple labels instead of uniquely MAs vs not MAs.

5.4 Results

Fig. 5.5 shows the results of the comparative FROC analysis on the Retinopathy Online Challenge data. Our technique compared very well with other submissions, particularly at a low false positive (FP) rate. This can be also appreciated in Table 2.3. Relatively high performance at a low FPs rate are particularly interesting for a screening setting, where it is not important to find all the MAs, but to find enough of them to decide that the patient needs referral.



(a) (b)

FIGURE 5.4: (a) Example of a retina fundus image with a MA magnified; (b) Inverted green channel of (a) with MAs automatically detected.

In the current unoptimized Matlab implementation, the whole analysis process takes ∼12 seconds per image on a 1.6 GHz machine with 4 GB of RAM. This figure can be easily reduced because around half of the total time is spent on the window alignment which is implemented iteratively. Also, the algorithm has the potential of greatly benefit of parallelization because of independent nature of the window based analysis.

We have also evaluated the performance of the algorithm on a synthetic model of a MA with variable dimensions, proportions and noise. The classifier is trained exactly on the training set used to evaluate the ROC images. These tests quantify the generalization ability of the algorithm on different MAs. The model f is based on a two dimensional normal distribution N with a maximum height of α and with a support window of 17×17 pixels. The choice of the model was dictated by the general consent among researchers that MAs mostly appear as 2-D Gaussians (Quellec et al., 2008).

$$N(x, \mu, \Sigma) = \frac{1}{2\pi|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \tag{5.3a}$$

$$f(x, \mu, \Sigma, \alpha) = \frac{max(N_{\mu,\Sigma})}{\alpha} \tag{5.3b}$$

where x is a two dimensional vector representing a coordinate in the window; μ is a two dimensional vector containing the coordinate at the centre of the window $[9\ 9]^t$; Σ is the 2×2 covariance matrix; $max(N_{\mu,\Sigma})$ is the maximum value for all the x for the given μ and Σ.

Fig. 5.6(a) shows the detection performance as a function of α with $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. The results are themselves normally distributed with a positive detection (i.e. MA prob.

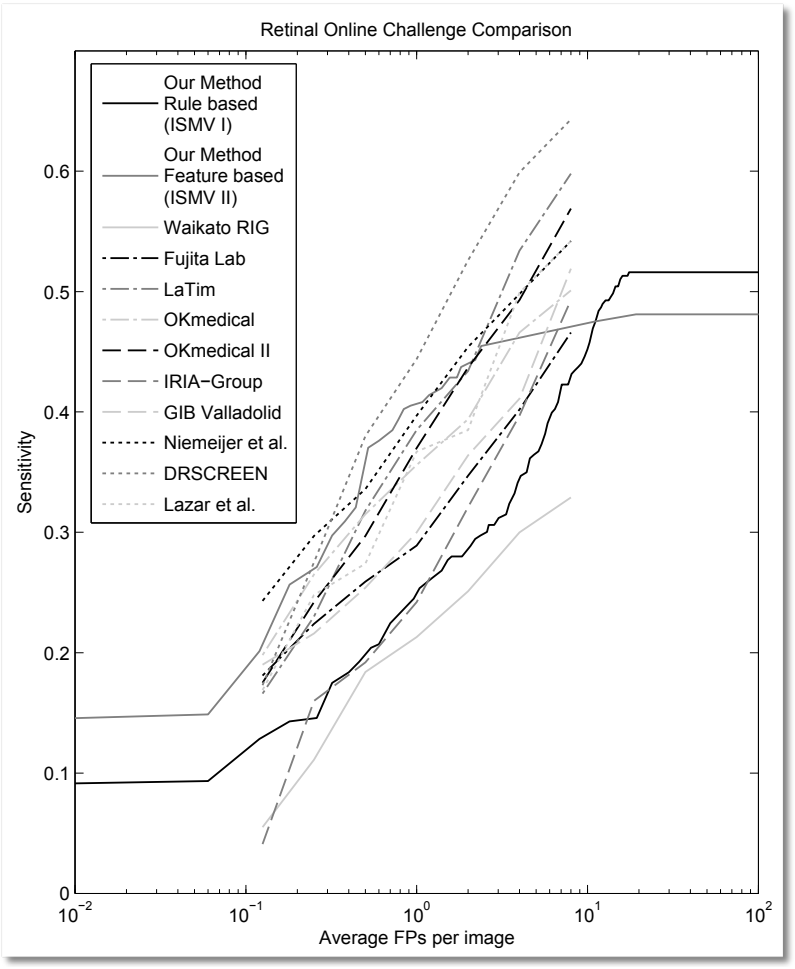


FIGURE 5.5: Comparison of the FROC curves of all the groups that participated at the ROC. On the y-axis, the average image based sensitivity is displayed, on the x-axis, the average number of false positives (FPs) found on each image is shown. Note that the FPs are plotted on a logarithmic scale.

≥ 0.5) with $0.045 \leq \alpha \leq 0.225$. Fig. 5.6(b) shows how the algorithm can detect circular objects with a great range of sizes. The size variation is simulated as a function of Σ with $\Sigma = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}$ and a fixed height ($\alpha = 0.16$). We obtained a positive detection for $0.3 \leq \sigma \leq 7.8$. Fig. 5.6(c) shows how the classifier can detect MAs that are not perfectly round, and when it stops doing so. We have simulated a round MA that becomes more and more elongated, up until it reaches a "vessel-like" appearance. This is achieved by using $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & \sigma \end{bmatrix}$, where σ goes from 0.1 to 50.

Fig. 5.6(d,e) shows various experiments for the detection of MAs in a noisy environment. In all cases the following parameters are used: $\alpha = 0.16$ and $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. On the

x-axis, the signal to noise ratio (SNR) measures the how much the original signal has been affected by noise:

$$SNR = \frac{(\sum f)^2}{(\sum f_{noisy} - \sum f)^2} \tag{5.4}$$

where $\sum f$ is the sum of all pixels of the MA model, and $\sum f_{noisy}$ is the sum of all pixels of the model affected by noise.

5.5 Discussion

In this chapter, we presented a MA detector based on a novel Radon-based approach. The Radon based features allow the detection of MAs directly on the original image without vessels segmentation. Also, they are inherently able to identify MAs of different sizes without multiscale analysis. Another advantage of the algorithm, is the ease of training. It does not require a large dataset, once some examples of MAs are shown to the classifier, it is simply a matter of dynamically selecting the negative examples on one or two images to make the algorithm "converge" to the desired performance. The technique presented has been evaluated on a publicly available dataset (where it ranked among the firsts) and on different synthetic MA models.

The substantial differences between the Radon-based and other authors' MA detection approaches is particularly promising for future developments, especially if combined with existing techniques. In fact, it has been shown that multiple detectors that measure uncorrelated features, generally achieve higher performances when combined together. The prime examples are the top performing algorithms in the ROC datasets by Antal et al. (2010); Niemeijer et al. (2005), that are combination of multiple MA detection techniques.

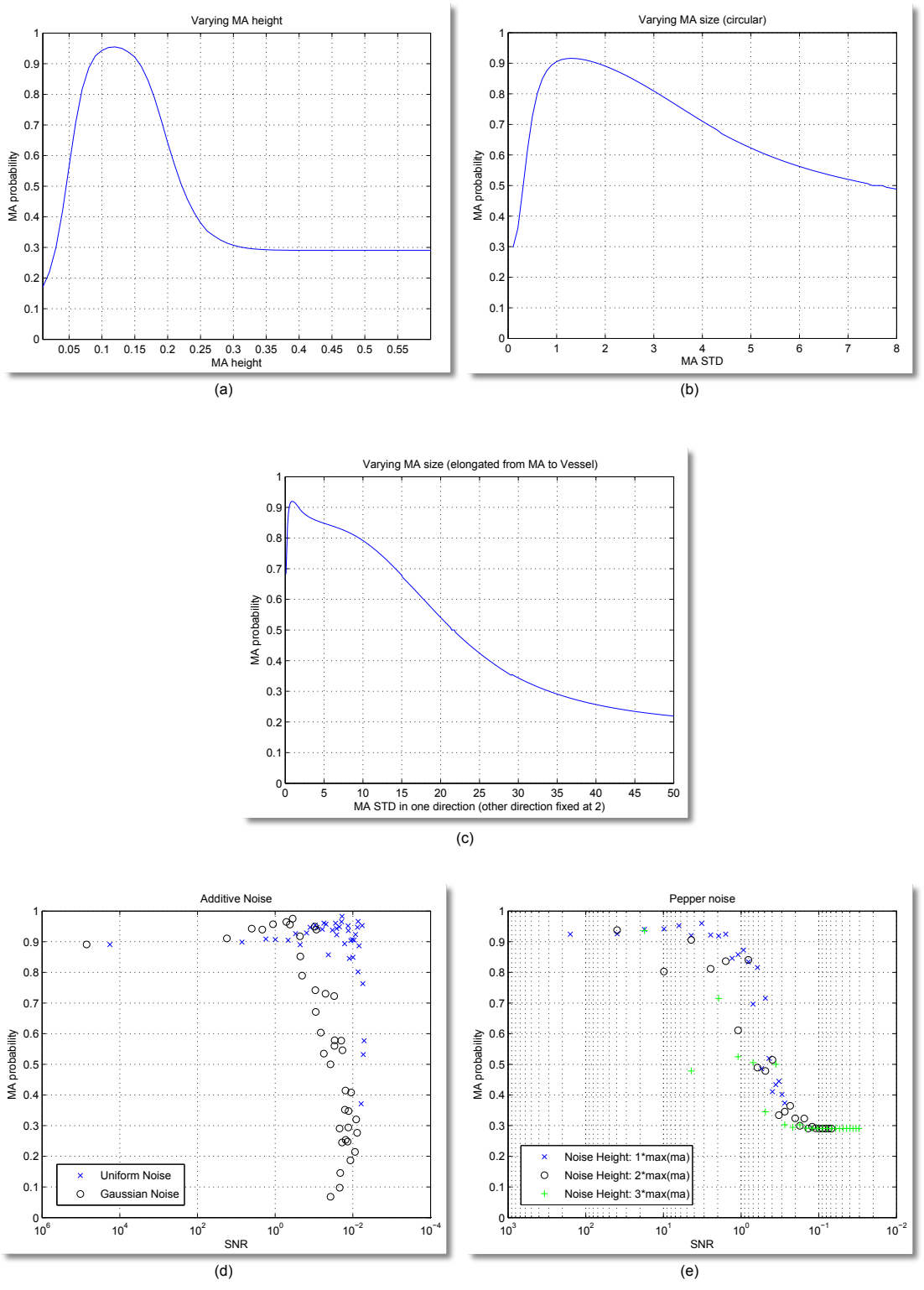


FIGURE 5.6: Tests on Synthetic Microaneurysms. (a) MA detection in function of MA height; (b) MA detection in function of size; (c) MA detection in function of not-proportional size changes (from circular MA to elongated vessel); (d) MA detection in function of two types of additive noise; (e) MA detection in function of pepper noise with different heights.

Chapter 6

Computer-aided Diagnosis

"Learn the changes, then forget them." - Charlie Parker

In this chapter, we present methods to automatically detect DME and DR by leveraging the lesion segmentation techniques introduced previously. For the evaluation of the algorithms three public datasets are used: the HEI-MED, MESSIDOR and DIARETDB1 dataset. Various tests are presented including cross-dataset testing (the classifier was tested on one dataset and trained on another one). While for the DR detection we statistically combine the output of the MAs segmentation, for the diagnosis of DME a novel set of features based on colour, wavelet decomposition and automatic lesion segmentation is employed. The techniques presented are quite computationally efficient, the total time required to obtain a complete diagnosis for DR and DME starting from an unseen image (including lesion, vessel and ON segmentation) is $\sim$22.3 seconds on a 2.6 GHz platform running an unoptimised Matlab implementation of the algorithms. Part of this work has been published in Giancardo et al. (2011c).

6.1 Introduction

In Section 2.4, the current state of the art automated systems for diabetic retinopathy screening and their potential impact on diabetic patients health were presented. In this chapter, we present our approach to the diagnosis of DR and DME. The medical conditions are diagnosed separately with two very different techniques, which both use a single macula centred colour fundus image captured at 45°. Even if such image is not suited for an in depth treatment plan by itself, according to Williams et al. (2004), it is enough for screening purposes. This consideration greatly simplifies the imaging process, in fact capturing different retina fields is not always a straightforward task, especially through an undilated pupil.

In Section 2.4.1, different grading scales for DR/DME were presented. Some of them are extremely detailed and permit to accurately monitor the progression of the disease. In the case of screening a yes/no referral decision is sufficient, as long as the positive decision is triggered when some form of DR is detected, even if minimal. Obviously, some cases can be more urgent than others and for this reason we have developed also an independent screening algorithm for DME only. DME is a complication of DR, and if found, it requires to be treated promptly, which is not always the case for mild forms of DR.

The method for DR diagnosis is heavily based on the output of the MAs detector presented in Chapter 5. The number of MAs detected and their probabilities are employed to generate a feature vector which is then classified to generate the probability of having DR. Even if other lesions typical of DR are not used for detection purposes, our methods achieve results comparable or better than other state of the art techniques.

The method for DME diagnosis employs the exudate segmentation algorithm presented in Chapter 4. However, this is not the only feature used. We present a classification framework that combines other types of analysis (based on colour and Wavelets) with the exudate segmentation map at image level. This is very different than typical approaches where each lesion is separately classified as a true positive or as false positive based on some measure. We create a fixed size feature vector regardless of the number of lesions with the great advantage of being able to train the algorithm without a dataset containing the ground truth at a lesion level, but only a single diagnosis per image. This is also possible thanks to our rule-based exudate segmentation algorithm.

The development of both algorithms was conducted on the HEI-MED dataset (and partially on the ROC dataset). None of the images present in the other two available public datasets, MESSIDOR and DIARETDB1, are used to tune the algorithms. They

are left entirely for testing purposes. We believe our experiments, as presented, show that our methods generalize well to other data sets which is a common problem in automatic retina image processing and machine learning in general.

We begin with a description of the datasets used in Section 6.2. Then, Section 6.3 summarizes the method used for the localization on the ON which was introduced by Tobin et al. (2007). Section 6.4 describes the techniques employed for the automated DME diagnosis. Section 6.5 presents the results for the DME diagnosis on the various datasets and the comparison with two retina experts. Section 6.6 describes the technique used for the automated DR diagnosis. Section 6.7 shows the results for the DR diagnosis with the MESSIDOR and HEI-MED dataset. Finally, the chapter concludes in Section 6.8 with some discussion of the results and approaches.

6.2 Datasets

In Section 4.2 the HEI-MED dataset was introduced. Besides this, the only two publicly available datasets that can be used to evaluate a DR/DME diagnosis algorithm are MESSIDOR and DIARETDB1. While the former provides the DR/DME diagnosis for each image only, the latter provides a lesion map generated by four different ophthalmologists but no explicit diagnosis. Hence, for the DIARETDB1 case, the DME diagnosis was generated by analysing the exudation ground truth map. If three out of four experts agreed that an exudate is present at any pixel, then the image is diagnosed with DME. However, it was decided not to use this dataset for DR diagnosis tests because of certain dark artefacts on the macula that strongly resemble MAs. They are probably due to some small round accumulation of dirt on the fundus camera lens. They are so similar to MAs that even for a human observer the only way to be aware of the problem is the fact that these artefacts always appear on exactly the same location. This is a problem not so uncommon in a real clinical setting, hence in Chapter 8 a technique to remove various types of lens artefacts by multiple fundus images is presented (which however concentrates on bright artefacts, which seem to be the most common). Table 6.1 shows some details of the datasets.

In the datasets used in this work, all the images were captured with a 45° FOV using different fundus cameras (even throughout the same dataset). In the MESSIDOR dataset, the images were obtained by a variety of research groups based in France. The heterogeneity, the substantial number of images, the public availability and the independence make these datasets ideal for an unbiased validation of our algorithms.

	HEI-MED	Messidor	DIARETDB1
DR Diagnosis			
Negative	103 (61%)	546 (45%)	n/a
Positive	66 (39%)	654 (55%)	n/a
DME Diagnosis			
Negative	115 (68%)	974 (81%)	51 (57%)
Positive	54 (32%)	226 (19%)	38 (43%)
ELVD Quality Metric			
Poor (ELVD < 0.5)	14 (8%)	62 (5%)	3 (3%)
Good ($0.5 \leq$ ELVD < 0.8)	31 (18%)	356 (30%)	27 (30%)
Excellent (ELVD ≥ 0.8)	124 (74%)	782 (65%)	59 (67%)
Ethnicity			
African American	104 (62%)	n/a	n/a
Caucasian	42 (25%)	n/a	88 (98%)
Hispanic	19 (11%)	n/a	n/a
Unknown	4 (2%)	n/a	1 (2%)

TABLE 6.1: Characteristics of the three datasets used in this chapter

6.3 ON Localization

The methods proposed for DR/DME diagnosis stem from the exudate and microaneurysms localization algorithms proposed in the previous chapters. One important step for the exudate segmentation algorithm is the localization of the ON centre. In the previous tests and experiments, its location has been taken from manually generated metadata, which allowed to evaluate the pure algorithm performance without being affected by this preprocessing step. However, for a fully automated diagnosis the ON location must be identified algorithmically. Section 2.2.4 provided an overview of current techniques for such task. For our purposes we have used the method introduced by Tobin et al. (2007).

The authors pose the ON localization as a Bayesian problem, where the prior probability $P(\omega_{ON})/P(\omega_{\neg ON})$ are respectively the 2-D likelihood maps of finding the ON at a given pixel and the likelihood of not finding it. Since all the fundus images that we employ are macula centred and have the same camera FOV (but different VFOVs and resolutions because they are acquired by multiple cameras) these priors are quite compact. In the final screening system all the images are vetted by a QA algorithm (like the one presented in Chapter 3), therefore we can assume a reasonable image alignment. The conditional densities to be combined to the priors to generate the posterior probabilities are $p(\nu|\omega_{ON})$ and $p(\nu|\omega_{\neg ON})$, where ν is a 4-dimensional vector containing vasculature derived features. The vasculature is represented as a binary map and it is automatically generated with the rule-based method by Zana and Klein (2001). The

binary vessel map is referred as I_b, its morphologically skeletonized version as I_{bs}. A description of the features used follows:

- **Retinal luminance**. This feature is the only one that is not based on the vascular structure. It is calculated with a M×N local window as the average luminance value on the green channel of the original image, where M = N = ON diameter. The ON diameter (ONd) is empirically estimated to be 130 pixels for an image 1024×1152. For images with a different resolution this measure is linearly scaled.
- **Vessel density**. The vessels tend to be denser the closer they are to the ON. This feature is captured by counting the average number of positive pixels in I_{bs} with a local window of $3 \cdot \text{ONd} \times 0.6 \cdot \text{ONd}$. By considering a much larger height than width for the local window, it is possible to capture the vertical vessels that are prominent nearby the ON.
- **Vessel thickness**. The average local vessel thickness is calculated by locally comparing I_b and I_{bs}. The local average in a window sized ONd × ONd is calculated for both images, then the local I_b average is divided by the local average value of I_{bs}.
- **Vessel Orientation**. This feature leverages the fact that the vessels close to the ON are vertical and then tend to a more horizontal direction the more they move away from the ON. The directionality is calculated with the steerable filter algorithm described by Freeman and Adelson (1991). The steerable filter used employs second derivative Gaussian combined with a Hilbert transform of this derivative.

The final location of the ON is calculated by finding the maximum in a confidence function R, which is calculated through the likelihood ratio of posterior probabilities as follows:

$$R = \frac{P(\omega_{ON}|\nu)}{P(\omega_{\neg ON}|\nu)} = \frac{p(\nu|\omega_{ON}) \cdot P(\omega_{ON})}{p(\nu|\omega_{\neg ON}) \cdot P(\omega_{\neg ON})} \tag{6.1}$$

6.4 DME Diagnosis - Methods

The method proposed for the DME diagnosis is based on the classification of single feature vector generated for each image. The feature vector is based on three types of analysis: the *Exudate probability map*, the *Colour Analysis* and the *Wavelet Analysis*. The rationale and approach for the feature computations is described in Section 6.4.4; which is followed by a description of the classifier used and the automatic techniques adopted to select the subsets of features that are employed in our experiments.

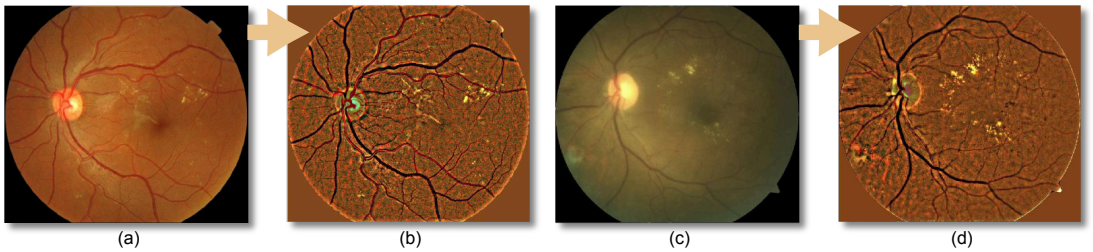


FIGURE 6.1: (a) Reference image I_{ref}; (b) Background-less reference image I^2_{ref}; (c) Image to be equalized I_{new}; (d) Equalized image I^4_{new}.

6.4.1 Exudate probability map

The exudate probability map is largely based on Kirsch's edges based detector presented in Chapter 4 which seems to show the best performance in the "naive DME diagnosis test" shown in Fig. 4.9. Assuming that the lesion segmentation output before thresholding is I_{kCand} (see Section 4.4.2.1), we scale it in order to create an exudate probability map I_{exProb} that represents $P(isExudate|I)$ as follows.

$$P(isExudate|I) = \begin{cases} 0, & \text{if } px < th_{low} \\ 1, & \text{if } px > th_{high} \\ \frac{px - th_{low}}{th_{high} - th_{low}}, & \text{otherwise} \end{cases} \tag{6.2}$$

where px is the pixel value of I_{kCand}; th_{low} and th_{high} are respectively: the lower boundary below which it is unlikely to identify any significant exudate, and the average upper boundary that can be identified in images with exudates. While th_{high} does not substantially influence the performance of the algorithm, th_{low} needs to be carefully selected in order to maintain most of the lesions without an excessive amount of false positives. Based on the exudate segmentation experiments shown in Fig. 4.7(a), we have picked a th_{low} of 4.5 that corresponds to a Sensitivity/Positive Predictive Value of 0.81/0.50 and a th_{high} of 30 which is where the FROC curve levels out. Note that the Sensitivity was calculated as the average ratio of lesions found in an image, i.e. a Sensitivity of 0.81 means that, on average, 81% of the lesions were identified on each image.

6.4.2 Colour Analysis

In Section 2.2.2, the fundus image normalization technique by Cree et al. (2005) was introduced. Their approach significantly reduces the inter-patient colour variability,

hence it makes possible the usage of colour as a reliable feature for classification purposes. The background is estimated by a large median filter, whose size is $\frac{1}{30}$ the height of the fundus image (no morphological reconstruction step is applied). The description of the process for a single colour plane follows.

$$\begin{aligned} I^2_{ref} &= I_{ref} - \text{medianFilter}(I_{ref}) \\ \mu_{ref} &= \text{mean}(I^2_{ref}) \\ \sigma_{ref} &= \text{std}(I^2_{ref}) \\ I^2_{new} &= I_{new} - \text{medianFilter}(I_{new}) \\ \mu_{new} &= \text{mean}(I^2_{new}) \\ \sigma_{new} &= \text{std}(I^2_{new}) \\ I^3_{new} &= (I^2_{new} - \mu_{new}) \div \sigma_{new} \\ I^4_{new} &= (I^3_{new} \times \sigma_{ref}) + \mu_{ref} \end{aligned} \tag{6.3}$$

where I_{ref} is the reference image, I_{new} is the image to be equalized and I^4_{new} is the equalized image. We have applied this process to the three planes of the RGB colour space using a single good quality reference image for all the datasets. With this equalization we aim to increase the reliability of colour based features which normally have a significant variability given the patient ethnicity, camera settings and image quality. Fig. 6.1 shows an example of the equalization of an image with a given reference image. Even if the two initial images have different ethnic backgrounds and quality levels (Fig. 6.1(c) is slightly blurred), the resulting images (Fig. 6.1(b,d)) have very similar colours, particularly with respect to areas of exudation.

Also, other unequalized colour features are evaluated with particular attention to colour spaces other than RGB, i.e. YCbCr and HSI.

6.4.3 Wavelet Analysis

In Chapter 4, we presented two new methods for the localization and segmentation of exudates. While the initial exudate probability map is derived from the Kirsch's edges based detector, the Wavelet-based detector seemed to offer good detection abilities and more importantly, is uncorrelated with other techniques with respect to the lesion properties detected. In fact, the former measured the contour edge strength of the lesions, and the latter took into account the edges inside the lesions. A combination between the two approaches seems like an obvious choice, however this integration is not done at a lesion level but rather at a global image level in this DME detection framework.

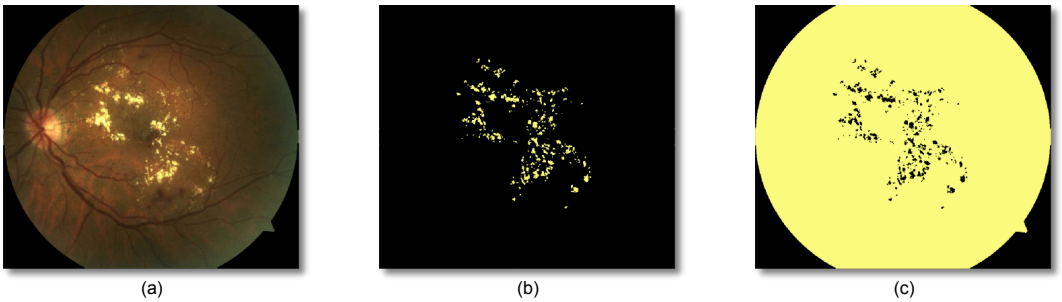


FIGURE 6.2: Inner and outer binary masks generated from the exudate probability map. In the binary masks the black colour corresponds to a pixel which is ignored, the yellow to a pixel which should be taken into consideration. (a) Original image; (b) Inner binary lesion map; (c) Outer binary lesion map.

Section 6.4.4 will explain how a single feature vector that represents the exudate probability map (therefore the Kirsch's edges based detector), the colour analysis and this Wavelet analysis is generated for each image.

This Wavelet analysis implies a process similar to the one previously described in Section 4.4.2.2. A stationary Wavelet analysis is performed up to the second level on I_i (I channel of the HSI colour space of the original image) with a Haar basis. Only the higher level details are kept and the process is inverted obtaining a regular image. Finally, we set 0 all the pixels below zero obtaining I_{wav} which shows a strong response in the exudative areas.

6.4.4 DME Feature Vector

wavMed, wavAvg, wavStd, wavMax, cbStd, cbMax, cbMin, crMed, crAvg,
crMin, crOutMed, rOutCreeAvg, gCreeStd, gOutCreeAvg, bCreeMax,
bOutCreeAvg, wavMedW, wavAvgW, wavStdW, wavMaxW, wavOutAvgW,
sMedW, sAvgW, sStdW, sMaxW, yStdW, cbMedW, cbAvgW, cbStdW,
cbMaxW, crMedW, crAvgW, crStdW, crMaxW, rCreeMedW, rCreeAvgW,
rCreeStdW, rCreeMaxW, rOutCreeAvgW, gCreeMedW, gCreeAvgW,
gCreeStdW, gCreeMaxW, gOutCreeAvgW, bCreeAvgW, bCreeStdW,
bCreeMaxW, bOutCreeAvgW
the feature names should be interpreted as follows:
<analysis plane><In/Out mask><statistics used> [W for Weighted]

TABLE 6.2: Feature set selected with Information Gain.

The selection of the appropriate feature vector to diagnose DME is a challenging problem. We want to characterize an image with a feature vector that has a fixed number of dimensions derived from a variable number of lesions, which might or might not be properly segmented. Other authors avoid this problem by providing a classification at

lesion level, where a set of features is generated for each lesion which is then classified as false or true positive. The diagnosis can be subsequently generated by combining the likelihood of the lesions. This approach requires a set of training images labelled at a lesion level, which implies more work to the reader who has to manually draw exudates and other lesions/pigmentation changes for each image. This might be troublesome for a classifier susceptible to outliers or when the samples in the dataset are not enough to average out the human error, which is inevitable since it is not a straightforward to identify all lesions (when there are many) and precisely define their edges. Instead, we are able to describe the lesions segmented as a whole by analysing the exudate probability map, the colour and the wavelet properties of the detected lesion set previously described. Two approaches have been examined, in the first one, the exudate probability map is converted to two binary masks which are overlaid on the colour and wavelet analysis outputs; in the second one, the exudate probability map is used to weigh the analysis outputs at a pixel level.

Fig. 6.2 shows an example of two binary masks, where one is generated by considering all the areas that have at least some chance of being exudative ($P(isExudate|I) > 0$), and the other is generated by considering the areas where $P(isExudate|I) = 0$. The masks are applied to the following image planes:

- *Wavelet:* The I_{wav} plane (see Section 6.4.3).
- *CreeRGB:* The three RGB channels after the retinal colour normalisation described in Section 6.4.2.
- *YCbCr:* The three channels of the YCbCr colour space.
- *HSI:* The saturation channel of the HSI colour space.

For each set of pixels extracted, the following statistics are calculated: mean, median, standard deviation, maximum and minimum. With these five measures we attempt to capture the correlation between the colour/wavelet analysis and the exudate probability map. This allows us to automatically identify problematic images which might have a high number of false positives because of consistent fibre layer reflection artefacts (which have a diffuse whitish colour and have different density in the wavelet analysis) or other unwanted false detections. By employing a mask where $P(isExudate|I) = 0$, we analyse also the areas not explicitly identified by the exudate segmentation. This appears to have a positive effect on the global classification, particularly on the images with a substandard exudate segmentation due to a noisy image or overexposed image.

The statistical measures described so far do not make full use of exudate probability map. To address this important aspect of the detection, we also use an approach where

the same binary masks, image planes and statistical measures are employed, but each pixel in the image planes previously described is weighted on $P(isExudate|I)$ before the statistics computation. This allows to include a measure of exudates segmentation uncertainty in the computation of the feature vector.

By combining the weighted and unweighted statistical measures, we obtain a total number of 80 features per image. The total number of features is pretty substantial with the risk of suffering of the "curse of dimensionality" during the classification phase. Hence, they underwent an automatic feature selection process. There are many techniques available to perform feature selection. We utilize *Information Gain*, an approach that seem to be relatively independent of the classifier used and that show reproducible results. The test has been run by splitting the HEI-MED datase into three folds. The feature selection is performed on each of the three subset. Each feature receives a final score that is the average between the three folds. This allows to select features that "overfit" the dataset, so that any test on the MESSIDOR or the DIARETDB1 dataset would not have any chance of being tainted by the feature selection process.

Information theory (Yang and Pedersen, 1997) provides us a straightforward way to apply conditional entropy to evaluate the significance of each feature (or attibute ($Attr$)) with *information gain*.

$$H(X) = -\sum_{x \in X} p(x)\ log_2[p(x)] \tag{6.4a}$$

$$H(Y|X) = \sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x)}{p(x,y)} \tag{6.4b}$$

$$IG(Attr, Class) = H(Class) - H(Class|Attr) \tag{6.4c}$$

where $H(X)$ is the entropy, $H(Y|X)$ is the conditional entropy, $IG(Y,X)$ is the information gain, $p(x)$ is the probability of x (i.e. the frequency) and $p(x,y)$ is the probability of simultaneously having x and y. We have employed this technique to reduce the full feature set (weighted and unweighted) to a number of 48 as shown in Table 6.2. This number was chosen by selecting only the features that had an average $IG > 0$.

6.4.5 DME Diagnosis Classification

In order to select the most appropriate classification strategy, we have run a series of tests with different combinations of classifiers/feature sets. The feature sets used are: all the unweighted features, all the weighted features, the combination of the two sets

and the features selected with Information Gain. As far as classifiers are concerned, we decide to cover the three different classification families described by Jain et al. (2000): *probabilistic*, *geometric* and *tree-based*. For the *probabilistic* family we tested the Naive Bayes classifier with two ways of estimating the prior probabilities, by assuming a Gaussian distribution of the data and by employing the Parzen Window approach. For the *geometric* family, two Support Vector Machines (SVMs) were tested, one with a linear kernel and one with the radial basis function as implemented in libSVM (Chang and Lin, 2001). For the *tree-based* family, the Random Forest algorithm with 10 trees was chosen (Breiman, 2001). In order to have a baseline, we reported the result of a nearest neighbour classifier which is likely the simplest classification method available (Duda et al., 2001).

The classifiers comparison tests are based on a Receiver Operating Characteristic (ROC) analysis with a Hold-One-Out (HOO) approach uniquely on the HEI-MED dataset. In HOO, the classifier is trained once for each image, each time holding out a different image which is used for testing. We want to avoid any type of optimization on the other two databases (Messidor and DIARETDB1) which will be used in the following section for an unbiased ROC analysis. Table 6.3 contains all the results for the classifiers comparison. The best Area Under the ROC Curve (AUC) for each feature set is highlighted in bold.

Classifier \ Feature Set	Unweighted	Weighted	Full	Information Gain
Nearest Neighbour	0.692	0.827	0.787	0.86
Naive Bayes (Gaussian)	0.78	0.837	0.864	0.897
Naive Bayes (Parzen Win.)	0.813	0.857	0.876	0.89
SVM (linear kernel)	0.914	**0.94**	**0.919**	**0.93**
SVM (radial kernel)	**0.923**	0.878	0.904	0.908
Random Forests	0.865	0.888	0.907	0.903

The results are expressed as AUC for the HOO tests on HEI-MED.

TABLE 6.3: DME classifier/feature selection tests with HEI-MED as testing target.

6.5 DME Diagnosis - Results

We have evaluated our computer-aided diagnosis system using three datasets. The results are presented as ROC curves where a positive image is an image showing signs of DME, and a negative image does not present any sign of this disease (but it might

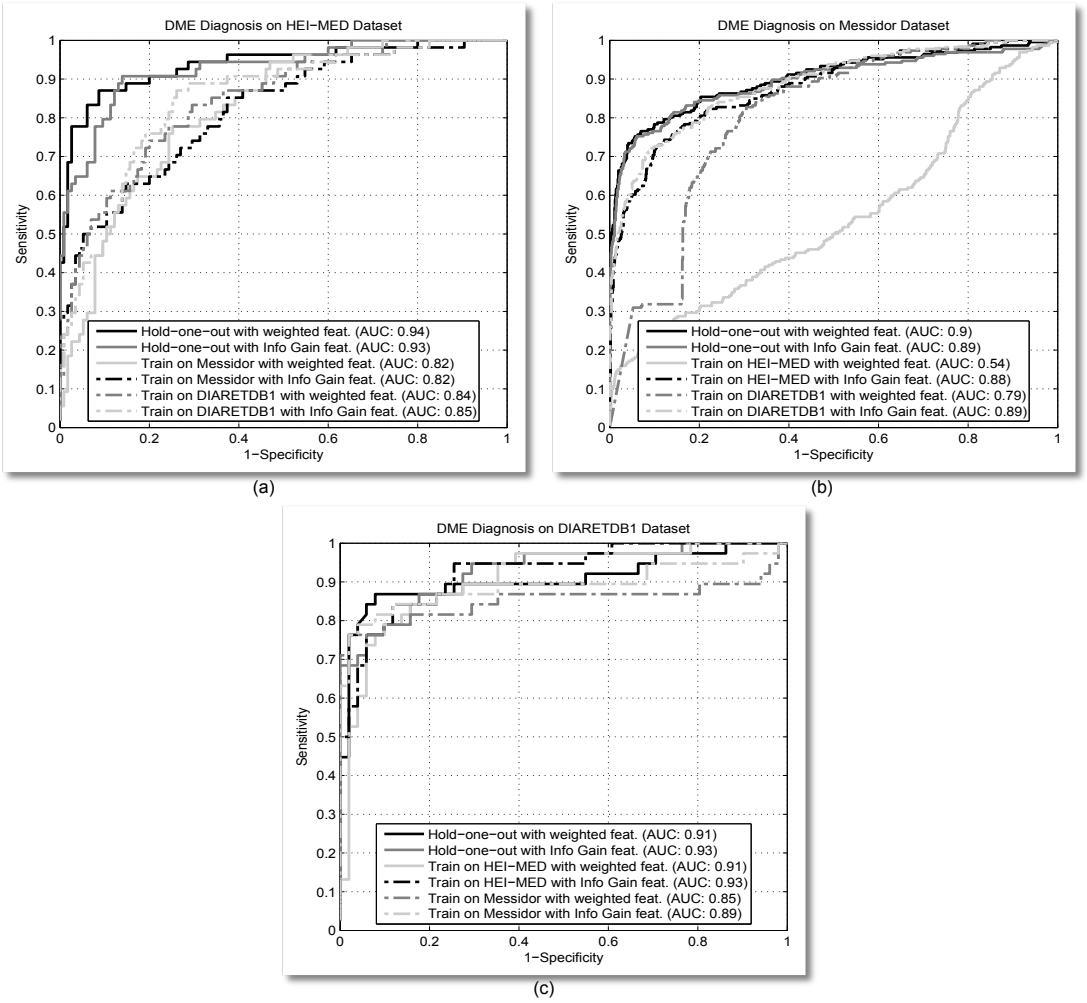


FIGURE 6.3: ROC curves for the DME diagnosis. Cross-datasets and HOO testing are employed on the three public datasets with a SVM (linear) classifier.

have other conditions). Based on the results of Table 6.3 we picked the SVM with a linear kernel since it showed the best performance for three out of the four feature set evaluated. The ROC curves are calculated by varying the threshold on the output positive diagnosis probability generated by the SVM classifier.

Fig. 6.3 shows the curves for the classification performed employing the two feature sets that had the best AUC in Table 6.3: the full unweighted set and the features selected with Information Gain (InfoGain). In each of the plots, three different tests are shown. Each test is represented by two ROC curves, one for each feature sets. The first test is a HOO, which should generally show the best performance since it is trained with images coming from the same dataset as the tested image. The other two are cross validation tests: the classifier is trained on each of the other dataset and tested on the current one. This type of test simulates very well the predicted performance of the

system on a real environment because it is trained on datasets with a different number of images, captured by a different camera/operator and labelled by a different expert than the current test set.

In Fig. 6.3(a) the performance on the HEI-MED dataset are shown. While the HOO tests reach considerable AUCs for both feature sets (0.94 and 0.93), the cross validation tests follow not so closely behind (although they are still acceptable). This suggests that HEI-MED dataset, albeit small, has a range of fundus appearance and conditions that might not be so well represented in the other datasets. This hypothesis is supported by the fact that in Fig. 6.3(b,c), the performance difference between HOO and cross validation tests is less significant (at least for the InfoGain features).

In Fig. 6.3(b) the performance on the Messidor dataset are shown. The InfoGain features perform consistently regardless of the type of test (AUC between 0.88 and 0.89). This is particularly encouraging in the validation of our technique because the HEI-MED and DIARETDB1 datasets have a significantly lower number of images than Messidor (169 and 89 vs. 1200). Interestingly, the weighted features have not performed as well: the AUC goes from 0.9 to 0.54. The reason is probably due to a subset of features that do not discriminate the same aspect of the disease in the different datasets. This stress the importance of feature selection and cross validation tests.

In Fig. 6.3(c) the performance on the DIARETDB1 dataset are shown. The performance of HOO and cross validation tests with the HEI-MED dataset are very satisfying, particularly with the InfoGain features (the AUC is 0.93 in both tests). The cross validation with the Messidor training is not as good. This is a constant throughout all the tests and it is probably due to some inconsistencies in the Messidor labelling which was confirmed by the two ophthalmologists. Fortunately, the Messidor dataset has a considerable amount of images, so we feel that these errors are averaged out. In addition, a subset of Messidor images were labelled by two retina specialists, who determined the presence or absence of exudates in order to diagnose DME on a random sample of 350 images of the MESSIDOR dataset, with 120 images exhibiting ME and 230 exhibiting no ME. We compared the performance of the automatic system by creating the ROC Curve and overlaid the Specificity/Sensitivity of the two retina specialists as shown in Fig. 6.4. The reference standard is provided by the MESSIDOR dataset. Two ROC curves are shown, both of them represents tests run on the 350 images only. One is a HOO test on these image, the other a cross validation test trained on the HEI-MED dataset. It can be seen that even if the experts beat the automatic system in an absolute sense, the system manages to obtain a comparable sensitivity.

The computational performance are evaluated on a Dual Core 2.6 GHz machine with 4 GB of RAM with an unoptimised Matlab implementation. The average time to generate

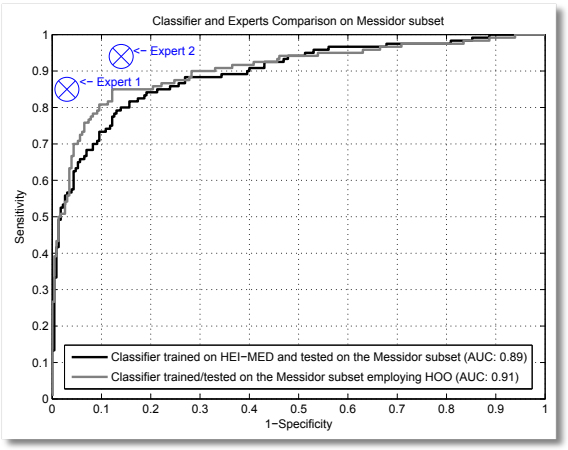


FIGURE 6.4: ROC curves for the DME diagnosis applied on the image subset diagnosed by the two experts. The feature set selected by Information Gain with a SVM (linear) classifier is used in both cases.

the exudate probability map is ∼1.9 seconds. ∼4.9 seconds needs to be added for the localization of the optic nerve. The average time to compute the other analysis types and classify the image is on average ∼2.5 seconds. Therefore, the total time to generate a DME diagnosis from a raw image (considering the optic nerve location) is ∼9.3 seconds. This time can be reduced further by optimizing the optic nerve localization algorithm, which does also calculate the macula location with some computationally expensive operations (Tobin et al., 2007).

	MESSIDOR	Expert 1	Expert 2	Algorithm
MESSIDOR	x	0.84	0.76	0.71
Expert 1	x	x	0.76	0.72
Expert 2	x	x	x	0.7
Algorithm	x	x	x	x

TABLE 6.4: Kappa-value Comparison.

	MESSIDOR	Expert 1	Expert 2	Algorithm
MESSIDOR	x	0.88	0.78	0.76
Expert 1	x	x	0.79	0.77
Expert 2	x	x	x	0.73
Algorithm	x	x	x	x

TABLE 6.5: AC1-statistics Comparison

To further quantify these tests, we used Kappa value/AC1-statistics as statistical concordance metrics (described in Section 2.2.1). In Table 6.4 and 6.5, we have compared the diagnosis between the experts, the MESSIDOR GT and the algorithm (trained on the HEI-MED dataset and tested on the Messidor subset). For the algorithm we have picked 0.22 as threshold, which simultaneously maximized both statistics.

6.6 DR Diagnosis - Methods

Our approach to DR diagnosis is entirely based on the output of the MA detector described in chapter 5. This is possible because the MAs are the very first manifestation of DR, hence if they are successfully detected it is possible to diagnose DR even in its mild form. The first DR classification strategy implemented is the enumeration of MAs with an estimated probability (P_{ma}) higher than 0.5. An image is deemed to have DR if more than a th number of MA are identified. By varying th, it is possible to perform a ROC analysis, which will be used as a baseline for the other tests. We note that this simple counting strategy does not make full use of MA probabilities. This information is pretty significant, judging the positive slope of the FROC curve obtained in the MA detector tests (Fig. 5.5). Therefore, we added a method that utilized the probability information by calculating the histogram of P_{ma} for all the MAs identified in an image. The histogram bins are used as a feature vector and then classified with a classic training/testing pattern recognition approach. The MAs considered for the histogram are only the ones with $P_{ma} > 0.5$. In Table 6.6, a variable number of histogram bins and classifier belonging to different families are evaluated on the HEI-MED dataset. It can be noted that many of the results are very close together, this is probably due the low dimensionality and compactness of the samples. The classifier/feature set employed in the testing phase have been highlighted in bold. During this phase it was noted that a some FP MAs were identified on the ON area. Since, we have the location of the ON already available it was decided to automatically remove the MA detected in ON area, which is estimated as 0.125 times the width and the height of effective retina area (i.e. VFOV). After the ON removal, the average performance improvement was estimated at $\sim$0.01 of AUC.

Classifier \ Feature Set	2 bins	3 bins	4 bins	5 bins
Nearest Neighbour	0.687	0.708	0.683	0.668
Naive Bayes (Gaussian)	0.809	0.819	0.824	0.818
Naive Bayes (Parzen Win.)	0.823	**0.829**	**0.833**	0.831
SVM (linear kernel)	0.825	**0.834**	**0.833**	0.823
SVM (radial kernel)	0.825	**0.831**	**0.833**	0.830
Random Forests	0.806	0.781	0.792	0.758

The results are expressed as AUC for the 3-fold tests on HEI-MED.

TABLE 6.6: DR classifier/feature selection tests with HEI-MED as testing target.

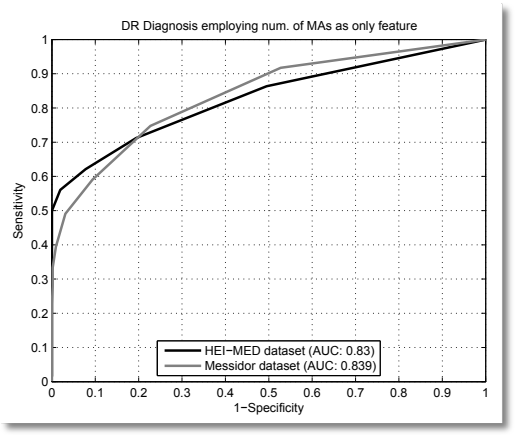


FIGURE 6.5: ROC curve for the DR diagnosis by counting the number of MAs only.

6.7 DR Diagnosis - Results

Fig. 6.5 shows the baseline results obtained by counting the MAs on each image. In both datasets the results are surprisingly good for a such simple classification technique. However, it should be remembered that the MA detector introduced is particularly competitive at a low FP rate which is required for an automatic screening system as this one.

Fig. 6.6 shows the tests performed with the best performing classifier/feature sets. In Fig. 6.6(a), the classifiers are trained on the Messidor and tested on HEI-MED. The best performance are obtained by the Naive Bayes classifier with Parzen window to estimate the samples distribution and a feature vector with 3 bins. The improvements in comparison with the baseline classification are not substantial, a 0.008 points of AUC. Fig. 6.6(b) shows the tests on the 1200 images of the Messidor dataset (with HEI-MED as training set). The best performing feature set/classifier are the same as the previous

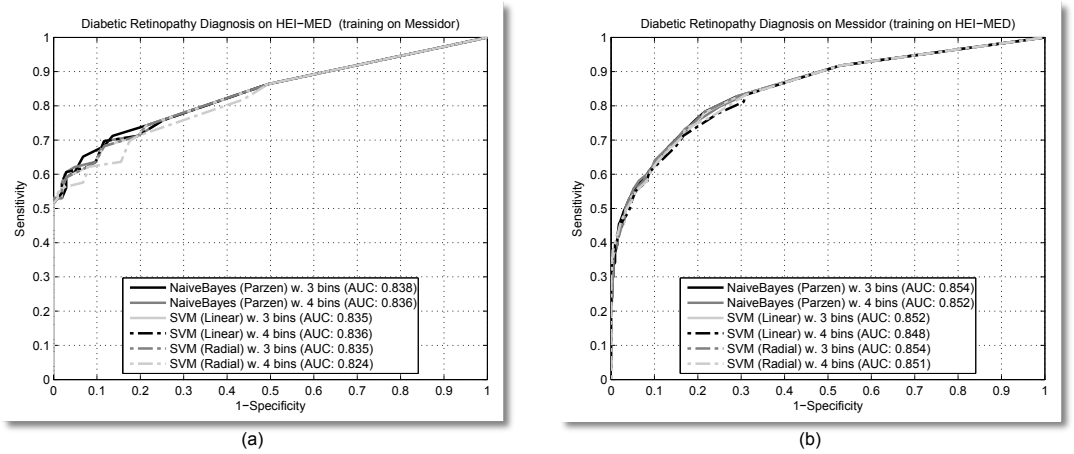


FIGURE 6.6: ROC curves for the DR diagnosis.

test. However, the improvement in comparison to the baseline is more pronounced: 0.13 AUC. This is even more significant considering the size of this dataset.

Agurto et al. (2010) are the only group that published the results of a DR screening algorithm employing the Messidor dataset, but they only used 400 of the images in the set (the ones labelled as *Lariboisière* in the original metadata). In their tests they obtained a AUC of 0.84, which is already lower in comparison to our results on the complete set (AUC of 0.854). For a fairer comparison, we tested our best performing classifier/feature set on the Lariboisière subset training it on the HEI-MED dataset. The result obtained improved even further, an AUC of 0.879 was achieved.

Being based almost entirely on the MAs segmentation, the time required by this classification technique is negligible is negligible (< 1 second) once the MAs are segmented.

6.8 Discussion

In this chapter, we have presented a new automatic system for the screening of DR and DME using non-stereo fundus images. Apart from their use of a previously detected lesion, the algorithms are very different between each other.

The DME screening method is based on an algorithm able to detect exudates with some attached confidence level without the use of machine learning methods to separate false positives from true positives, on a colour space analysis and on new methods to characterize the lesions by the means of wavelet analysis. To our knowledge, our approach for the creation of the feature vector with a inner and outer lesion maps has never been

attempted before and proved to be quite successful. We are confident that this type of approach can be applied to all the problem domains where a diagnosis (or other types of classifications) needs to be performed on the basis of an uncertain lesions (or other defects) segmentation. The DR screening method is much more compact because it leverages the good performance of the MA detector algorithm.

We have tested our algorithms against three different datasets. For DME we have achieved an AUC between 0.88 and 0.93. We compared this results with the performance of two retina specialist obtaining comparable results with different test modalities (ROC, K-value and AC1-statistic). A system to diagnose diabetic retinopathy that was tested with the largest dataset (approximately 15 000 patients) is presented by Niemeijer et al. (2009a). They report an AUC of 0.88. In another recent study, Agurto et al. (2010) diagnose diabetic retinopathy avoiding the usage of manually segmented lesion too, by employing AM-FM features. They have obtained an AUC of 0.84 by considering a subset of 400 images of the MESSIDOR dataset. Our DR detection technique reaches an AUC of 0.879 employing the same 400 images.

The total time required to generate a complete screening diagnosis (from an unseen image) is $\sim$22 seconds per image on a 1.6 GHz machine with 4 GB of RAM in a Matlab implementation with each component run sequentially. By running DR and DME in parallel it is expected to achieve $\sim$17 seconds of running time.

In order to effectively test the feasibility of these techniques in an environment as the one of healthcare, tests with datasets much larger than the ones used need to be performed and compared with the performance of retina experts. However, it is felt the usage of three independent (and public) datasets made a strong point about the effectiveness of the algorithms.

Chapter 7

Macula Swelling Detection with Multiple Images

"There are no wrong notes." - Thelonious Monk

In this chapter, we propose a novel technique that uses uncalibrated multiple-view fundus images to analyse the swelling of the macula. This innovation enables the detection and quantitative measurement of swollen areas by remote ophthalmologists with inexpensive fundus cameras. This capability is not available with a single image and prone to error with stereo fundus cameras. We also present automatic algorithms to measure features from the reconstructed image which are useful in the automated diagnosis of early macular edema, e.g., before the appearance of exudation. The technique presented is divided into three parts: first, a pre-processing technique simultaneously enhances the dark micro-structures of the macula and equalises the image; second, all available views are registered using non-morphological sparse features; finally, a dense pyramidal optical flow is calculated for all the images and statistically combined to build a naive-height-map of the macula. Results are presented on three sets of synthetic images and two sets of real world images. These preliminary tests show the ability to infer a minimum swelling of 300 microns and to correlate the reconstruction with the swollen location. Part of this work has been published in Giancardo et al. (2011e).

7.1 Introduction

The ease of use, relatively low cost, and utility of fundus cameras in detecting various retina diseases have made possible their widespread use in clinical practises. However, there are many cases when the 3D structure of the retina would be useful or even essential for a correct diagnosis. This is especially the case for the initial stages of Diabetic Macular Edema (DME), a common cause of vision loss and blindness, as discussed in Section 2.1.4.2. DME occurs in the retina of diabetic patients due to leakage of fluid within the macula which creates diffuse swelling and cystic changes that are not immediately visible with monocular fundus images. The classical way to detect this fluid is either by waiting for its absorption which turns into exudation or by analysing the 3D shape of the outer layer of the retina (with particular focus in the macula area).

Typically, the 3D information is obtained with stereoscopic fundus cameras or Optical Coherence Tomography (OCT) instruments. However, stereo glasses are required to visualize the depth field in commercially available stereoscopic fundus cameras. Hence, they do not extract numerical information that can be used as features in automatic methods and are prone to substantial intra/inter-reader variability on the diagnosis (Li et al., 2010). This is a serious drawback for their use in automatic diagnosis of macula swelling. The ideal solution, a OCT system in conjunction with a fundus camera, has considerable cost disadvantages and requires additional training for the operators who will be capturing the images. This seriously limits its practicality in a screening environment. Thus, an algorithm able to infer the 3D shape of the retina from multiple fundus cameras images would be extremely beneficial in terms of cost, and also can add capabilities for automatic diagnosis of fluid leakage in the macula area which can be invisible in standard fundus images (see Fig. 7.6a). We note that obtaining multiple images is rather simple with modern high-quality, easy-to-use general-purpose fundus cameras, as the image acquisition time is fast (a few seconds for each) and painless for the patient.

In the literature there are examples of 3D retina reconstruction algorithms that concentrate mainly on optic nerve (ON) reconstruction for the detection of glaucoma, either with a calibrated stereoscopic fundus camera (Corona et al., 2002; Xu and Chutatape, 2006; Nakagawa et al., 2008) or with multiple views from a standard monocular fundus camera (Chanwimaluang et al., 2009; Liu et al., 2008), provided that the aperture of the monocular camera is kept smaller than the eye pupil (Martinello et al., 2007). Choe et al. (2006) present a reconstruction technique on the entire FOV of the image employing fluorescein images which require a special filter on the fundus camera and the injection of a contrast agent, a technique that is too invasive for screening purposes and that cannot be performed outside a specialised clinic due to possible unwanted secondary effects.

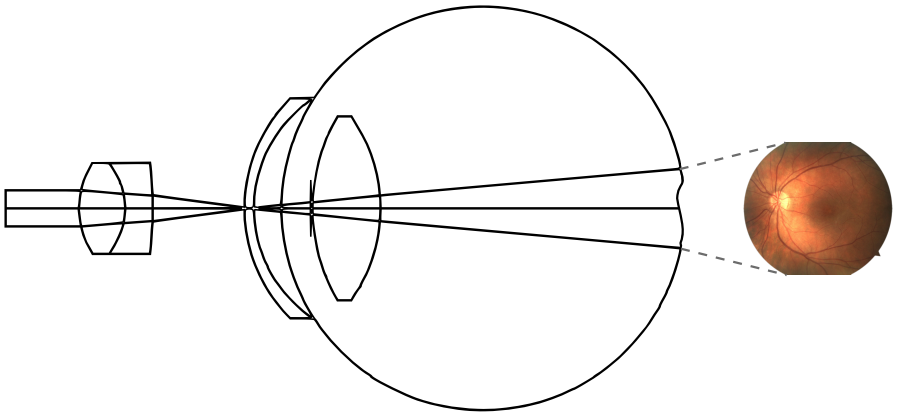


FIGURE 7.1: Schematic lateral view of the virtual fundus camera.

However, to our knowledge, no research group has proposed any method to directly identify swelling of the macula (an almost texture-less area of the retina) with fundus images. The lack of a fixed multi-view configuration for the PoC fundus cameras and the impracticality of a calibration procedure are a difficult but common problem in 3-D multi-view geometry. There are various approaches that are able to successfully tackle this problem, such as bundle adjustment/Structure from Motion (SfM) (Hartley and Zisserman, 2003) or Simultaneous Localisation and Mapping (SLAM) (Durrant-Whyte and Bailey, 2006). All these techniques require strong salient point correspondences to simultaneously estimate the camera pose and the 3-D structure. This is not possible in the macula which has little texture and where "corners" cannot be reliably tracked and matched.

We propose a novel algorithm able to identify macular swelling through the reconstruction of a naive-height-map of the macula area from multiple fundus images with an unknown translation (roughly parallel to the eye), captured by an uncalibrated fundus camera. In our experiments, we show how retina "blisters" can be identified even in areas where there is no apparent texture visible using four fundus images.

In Section 7.2 we explain how the eye/camera model is built and how the real patient images were acquired; Section 7.3 introduces the details of the algorithm; then, Section 7.4 shows the results for the model and the real images; finally, Section 7.5 discusses the implications of this technique.

7.2 Materials

The development and testing of our algorithms requires a gold standard for comparison with the reconstruction (i.e. a profile from an OCT instrument) and images of patients with blisters of different heights. The collection of this type of data is a challenging task by itself; hence to speed up the development phase and to improve the evaluation we have developed and tested the algorithm with a faithful ray-traced model of the camera-eye-retina system. We have made this model and related software available online (`http://vibot.u-bourgogne.fr/luca/`). The optical model was initially designed in Zemax© to create a credible representation of fundus camera/optical eye structure and then modelled in POVray, a freely available ray-tracing tool. In our model, we place a real image of the retina fundus at the back of the eye and extrude it with a given height map that acts as a ground truth. Fig. 7.1 shows a lateral view of the model. We acquired four virtual shots by laterally translating the camera at 10°, 5°, -5°and -10°(the angle is calculated between the centres of the objective lens and the cornea). Fig. 7.2(a-d) show examples of the images obtained by translating the virtual camera. The resolution used is 1024 x 768.

In order to acquire preliminary data on the clinical viability of the method, we asked a camera operator to obtain four images with the only constraint being lateral translation between the images. The operator acquired two sets of images from two patients with associated OCT data for verification. One patient shows a blister in the OCT image which is invisible to the fundus camera (Fig. 7.6(a)) while the other has a completely healthy macula (Fig. 7.7(a)). The operator translated the camera a considerably smaller amount than the translation present in the simulated images (estimated at 10%) which reduces the baseline and makes the reconstruction more challenging, but provided an interesting testbed for these preliminary experiments.

7.3 Method

The reconstruction algorithm is divided into three phases: *pre-processing*, *rigid registration* and *naive-height-map reconstruction*. First, the fundus images are enhanced to remove background information and reflections from the nerve fibre layer (NFL). This step also enhances the micro structures in the macula area. Second, all images are rigidly registered by employing fiducial points independent of retina morphology. Finally, the naive-height-map is reconstructed by exploiting the statistical distribution of a dense optical flow analysis between images.

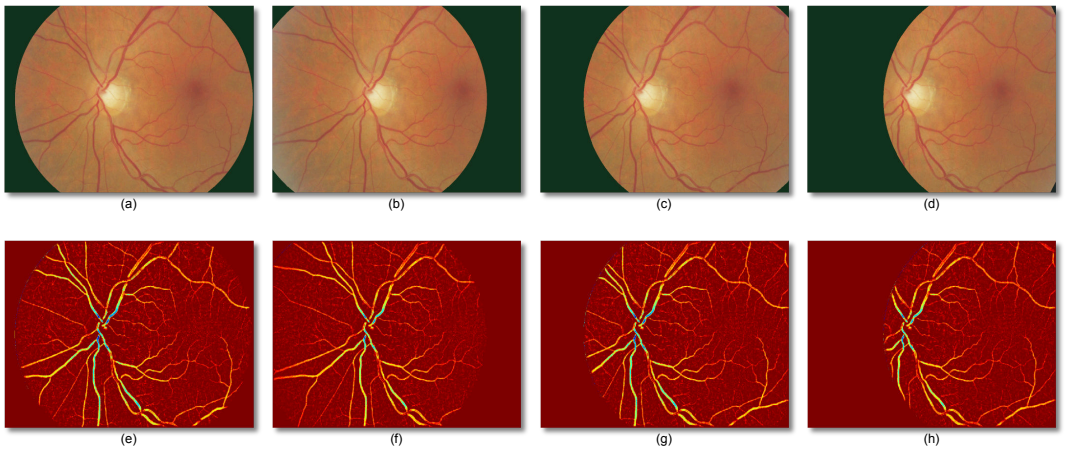


FIGURE 7.2: (a)-(d) images captured with the virtual fundus camera; (e)-(h) images after the pre-processing phase.

7.3.1 Pre-processing

We start the pre-processing by extracting the green channel I_g from each image and performing an initial estimation of the background by means of a large square median filter, whose size is $\frac{1}{30}$ the vertical size of the fundus image. This estimation is enhanced with the addition of the morphological reconstruction step described in Section 4.4.1. The estimated background is subtracted from the original image with 16-bit signed precision to maintain negative pixel values. The image obtained shows a distinct grey level distribution: the highest peak of the histogram is always centred on zero regardless of the ethnicity of the patient, disease or point of view of the camera. We are able to obtain the texture enhanced version of the macula as shown in Fig. 7.2 by maintaining the absolute value of the negative pixel values only. With this technique, we are able to remove most of the illuminations changes due to the different positions of the camera.

7.3.2 Rigid Registration

The motion between fundus images is compensated through a registration technique inspired by Cattin et al. (2006). By using Speeded Up Robust Features (SURF) (Bay et al., 2008), a local descriptor that can quickly generate an informative 64-dimensional vector for a given point, we selected the salient points with a SURF quality value greater than 0.0001. Then, all salient points are matched throughout all views employing the Euclidean distance between SURF vectors. Only the points with a distance ratio $\frac{best\ match}{second\ best\ match}$ larger than 0.7 are kept. The RANSAC algorithm with rotational homography is used to estimate the outliers (i.e. the points whose geometry is unfit for

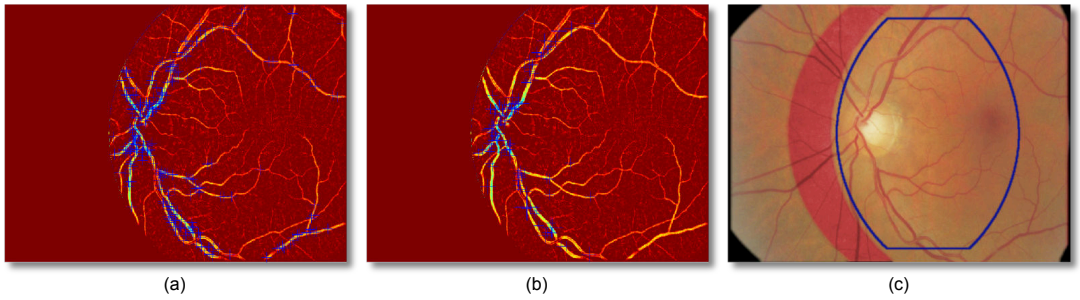


FIGURE 7.3: (a) initial SURF features extracted; (b) SURF features employed for the registration; (c) four registered images superimposed over each others. The colour of an image has been modified to show the registration junctions. The blue contour shows the common VFOV detected.

the rotational rigid transformation) (Hartley and Zisserman, 2003). The area of interest is calculated with a binary AND operation on all the VFOV, which are detected with method described in Section 3.2.1.1. Fig. 7.3 shows an example of the rigid registration.

7.3.3 Naive-height-map Reconstruction

The reconstruction of the macular shape is based on the fact that most of the retinal structures have a very small height (less than 0.8mm) in comparison to the planar size of the retina; hence, the rigid registration allows a nearly perfect alignment of most of the areas apart from the ones closer to the camera, i.e. blisters. We exploit this characteristic by building $n-1$ disparity maps, where n is the number of images and each map is built using a common reference image identified as im_0, i.e. $im_0 \leftrightarrow \langle im_1 \dots im_{n-1} \rangle$. Because of the lack of strong salient points in the macula, the disparity maps are approximated by computing the dense optical flow with the pyramidal Lucas-Kanade method (Bouguet, 2000), for each pixel of the image (in the common VFOV) with a window of 19 pixels and 3 pyramids. The magnitude of each vector is computed in order to obtain the provisional disparity maps $\langle im_1^{dmap} \dots im_{n-1}^{dmap} \rangle$.

Assuming that in the areas where there is no texture the flow vector obtained will be random while in the areas with little texture the noise will be normally distributed, we can compute a final naive-height-map im^{nhm} as follows:

$$im^{nhm} = \frac{1}{n-1} \sum_{i=1}^{n-1} im_i^{dmap} \tag{7.1}$$

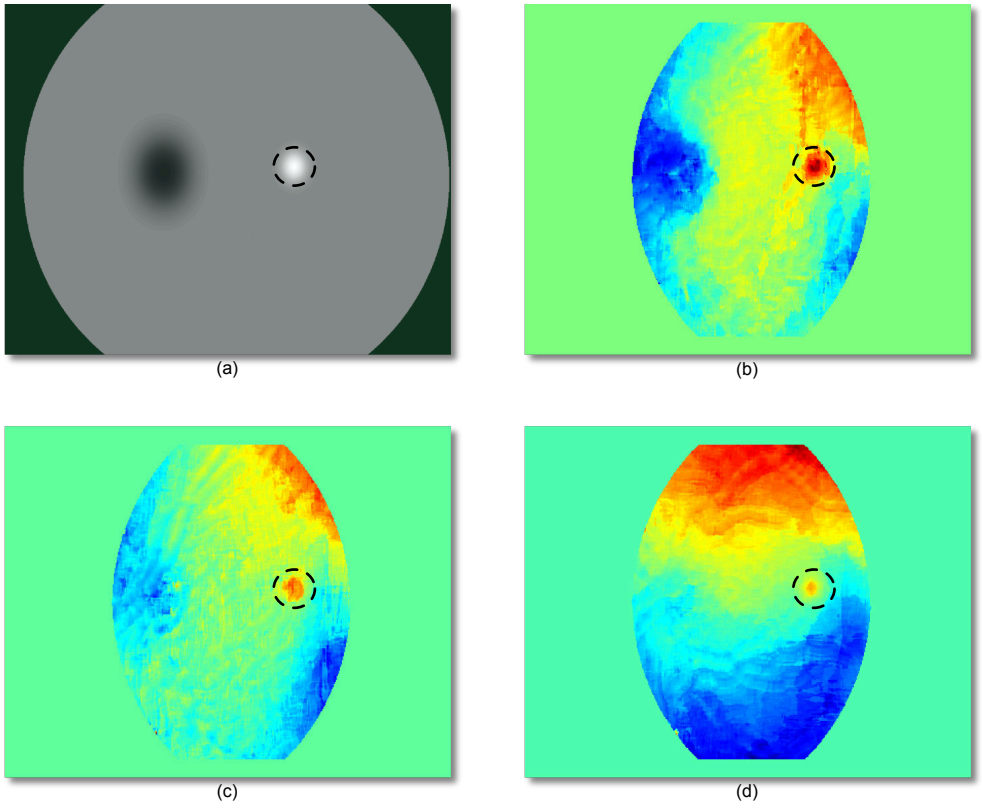


FIGURE 7.4: (a) synthetic height map used to generate the retina 3D structure (the white spot represents the blister); (b)-(d) reconstructed naive-height-maps containing a blisters of decreasing height, respectively 0.5mm, 0.4mm and 0.3mm. The area contained by the dashed circle shows the position of the original blister.

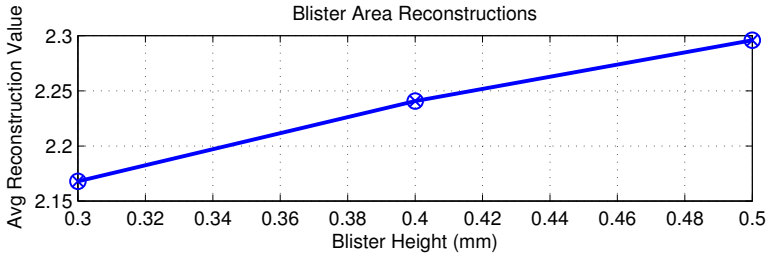


FIGURE 7.5: Mean value of the reconstructed naive-height-map in the area shown in Fig. 7.4 for the different heights of the blisters generated with the virtual camera.

7.4 Results

Using the images generated with the virtual camera, it was possible to reconstruct the exact position of the blister in the macular area as shown in Fig. 7.4. In Fig. 7.5, we show the reconstruction ability numerically by calculating the mean of the naive-height-map

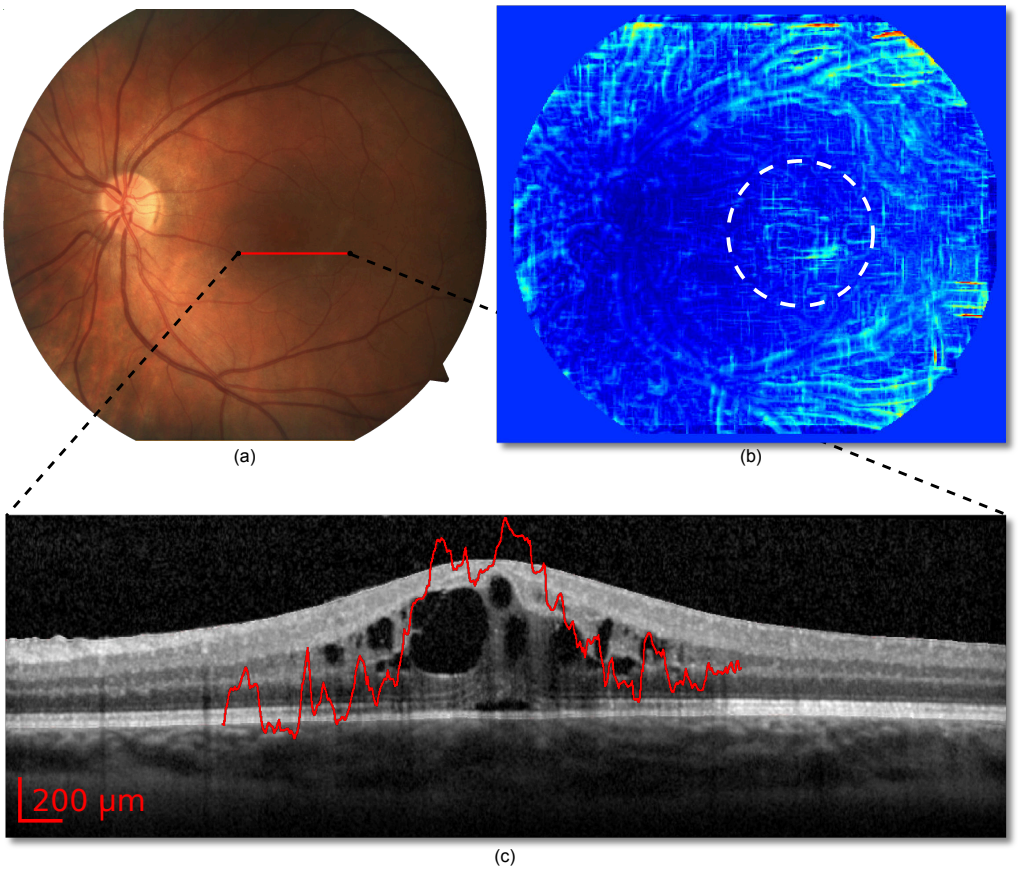


FIGURE 7.6: Patient A; (a) real fundus image of a patient with a blister invisible to a standard fundus camera; (b) reconstructed map showing the blister, the white circle shows the area of interest; (c) OCT image of the highlighted area with a red overlay of the 3D reconstruction obtained.

value in the same area. In the experiment, this value linearly correlates with the height of the blister.

Figure 7.6 and 7.7 show the algorithm's reconstruction ability on the images acquired by the operator as described in Section 7.2. Based on the fundus images alone, these patients do not exhibit any abnormality. In Fig. 7.6(b) and Fig. 7.7(b) we show the reconstructed naive-height-maps, and in Fig. 7.6(c) and Fig. 7.7(c) we show the three-dimensional overlay of the reconstruction with the accompanying OCT. Although the reconstruction has a lower quality than the simulated images of Fig. 7.4, the healthy macula is clearly discernible from the swollen one. The 3D reconstruction overlay (Fig. 7.6(c)) demonstrates a strong correlation with the actual retina edema seen on OCT imaging. The difference in average height in the macula area between the healthy and non-healthy image is 2.13, a result that (if confirmed by further experiments) will be a clinically significant value for the automated detection of macular swelling. We were

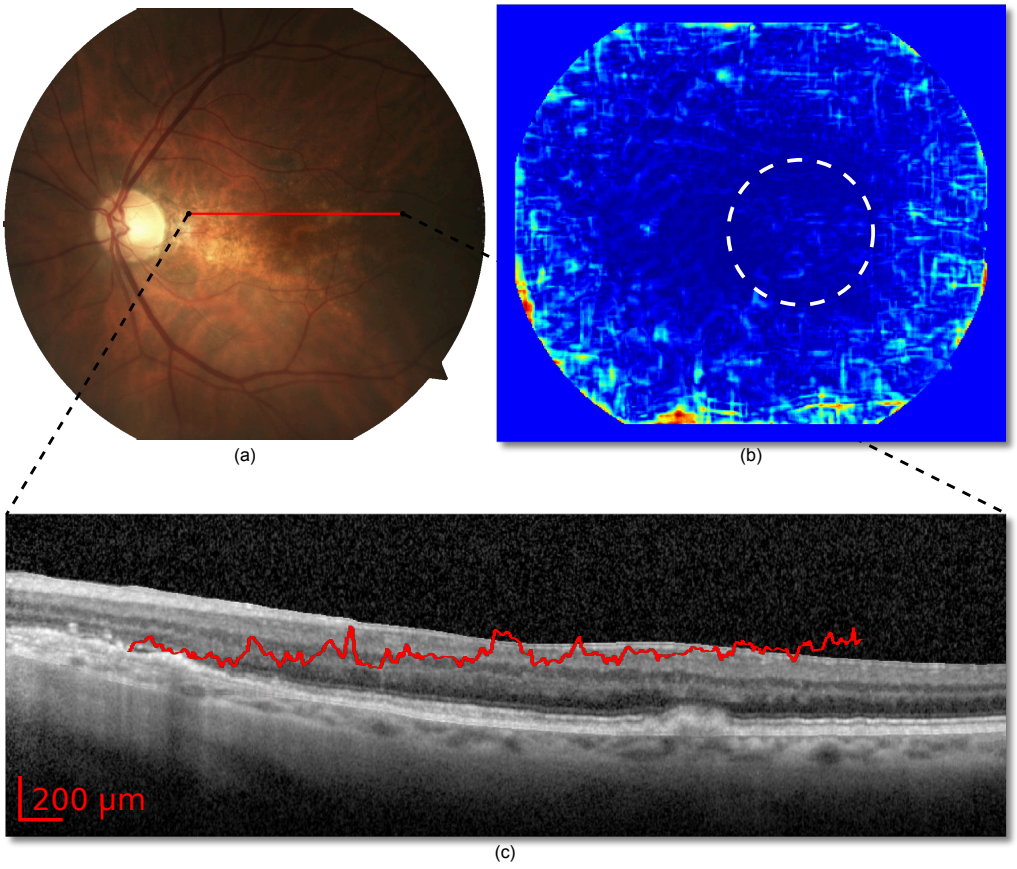


FIGURE 7.7: Patient B; (a) real fundus image of a patient with an healthy macula; (b) reconstructed map, the white circle shows the area of interest; (c) OCT image of the highlighted area with a red overlay of the 3D reconstruction obtained.

able to analyse a set of 4 images in under 5 minutes on a 2.2GHz machine with 4GB of memory using a Matlab algorithm implementation.

In the reconstructions we notice artefacts of increasing strength starting roughly outside the macula area. They are due to lens distortions and miscalculation of the optical flow; however, these are not a problem because our algorithm is focused on the macular area which can be automatically identified by other image processing algorithms (see Section 2.2.4).

7.5 Discussion

Fundus cameras are an effective tool for diagnosis of retina diseases but cannot reliably detect depth which is a key indicator of the early phases of diseases such as DME. Multi-view image reconstruction can be used to perform 3D reconstructions, but the

lack of strong salient feature points in the macula is a potential flaw in many existing multi-view reconstruction approaches. In this chapter we have presented a method which performs a straightforward statistical analysis of noisy disparity maps to generate a naive-height-map that can be used to display and potentially measure blisters in the macular area. We have shown its feasibility on synthetic images constructed using a computer model and have also illustrated its effectiveness on two data sets of four fundus images and associated OCT data. The proposed algorithm is a promising addition to automated image analysis using inexpensive fundus cameras well-suited to retina screening applications.

Chapter 8

Quality Enhancement with Multiple Images

"Anche se voi vi credete assolti siete lo stesso coinvolti." - Canzone del Maggio, Fabrizio De André

In this chapter, we propose a technique that employs multiple fundus images to obtain a single higher quality image without nerve fibre layer (NFL) or camera lens related reflections artefacts, which also compensates for a suboptimal illumination. The removal of bright artefacts, can have great benefits for the reduction of false positives in the detection of retinal lesions by automatic systems or manual inspection. The fundus images are acquired by changing the stare point of the patient but keeping the camera fixed. Between each shot, the apparent shape and position of all the retinal structures that do not exhibit isotropic reflectance (e.g. bright artefacts) change. This physical effect is exploited by the algorithm. Part of this work has been published in Giancardo et al. (2011a).

8.1 Introduction

In this chapter, the early results of a novel post-processing technique able to remove the reflectance artefacts due to the nerve fibre layer (NFL) or dirty camera lenses are presented. This method can also automatically identify the under/over-illuminated areas and compensate by substituting them with the redundant information found on the other images (if available). This is made possible by capturing redundant images of the same subject from slightly different angles. This has the effect of changing the characteristics of the reflectance artefacts.

While various groups have built algorithms to automatically evaluate the quality of fundus images (see Chapter 3 and Section 2.2.5), to our knowledge, little or no work has been published to increase the retina image quality after the acquisition. Various techniques to equalise the image were introduced in Section 2.2.2, but no attempts have been made in the removal of NFL artefacts or the exploitation of multiple images of the same patient. This is partially because this type of reflectance artefacts have become much more apparent only in the recent years, with fundus cameras that do not require the dilation of the pupil by use of mydriatic eye drops and the increasingly young age of the patients to be screened. In fact, non-mydriatic fundus cameras require a high power flash, in order to allow enough light to enter the pupil which is not completely dilated. This increases the chances of reflectance artefacts of any kind, particularly in the areas where the NFL is thick. This undesirable effect is more evident on the young patients because the NFL thickness (hence reflectance) decreases with ageing.

On the other hand, the image quality enhancement based on redundant images of the same subject is not new. Many authors have developed *super-resolution* algorithms to create a single high resolution image from multiple low resolution ones Farsiu et al. (2006); Park et al. (2003). Unfortunately, these techniques are very specific in improving the quality in a resolution sense, and while they can deal with blur and salt and pepper noise, they do not seem to be suited for the removal of large structures such as NFL reflections.

In Section 8.2, we start with a description of the protocol used to capture the retina images and the visual effects achieved; Section 8.3 describes the algorithm; Section 8.4 presents the results obtained and Section 8.5 discusses them.

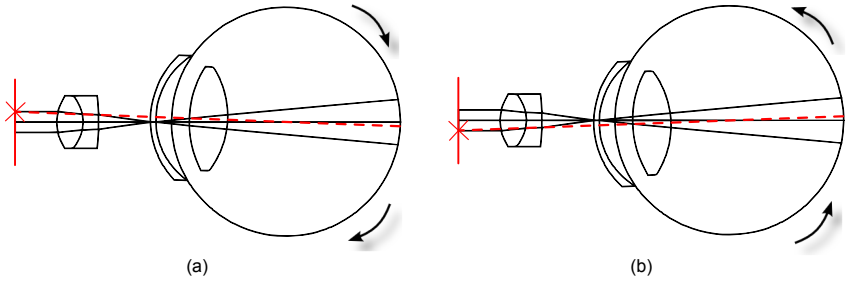


FIGURE 8.1: Examples of image acquisition with Fundus camera/eye top views. The red X represents the fixation target, the dotted red line the ray of light going to the fovea (the area that we employ when we look at an object). (a) Left shift of the fixation point. (b) Right shift of the fixation point.

8.2 Image Acquisition

One characteristic that differentiates a NFL from the rest of the retina is its physical reflectance properties. The NFL shows a considerable specular reflectance when interacting with the camera flash wavefront, while the rest of the retina shows a largely diffuse reflectivity behaviour. This means that the energy of two rays of light hitting the same area of the NFL will be almost completely reflected to two different directions, whenever the angle of incidence differs. The visible effect on the image sensor is a bright artefact that appears or disappears depending on the angle of incidence of the light. This effect is not noticeable in the other areas of the retina which show a diffuse reflectivity behaviour, where the light energy is largely reflected in all directions (with reduced energy) regardless of the angle of incidence of the rays.

Changing the incidence of the light is not a trivial task for fundus imaging, as the surface of the retina imaged is limited (∼7mm of diameter in our set-up) and all the light has to go through the pupil as shown in Fig. 8.1. Fortunately, modern fundus cameras have a manually changeable fixation point which allows a natural alignment of retina whenever the patient stares at it as shown in Fig. 8.1(a-b). If we change the position of the fixation point the angle of incidence of the wavefront coming from the flash is changed, therefore the effect previously described can be easily achieved with any fundus camera that allows the repositioning of the fixation target.

In our experiment we asked an inexperienced operator to randomly place the fixation target and to capture seven images without worry about the quality. These images are shown in Fig. 8.5, where it is possible to notice the appearance and disappearance of

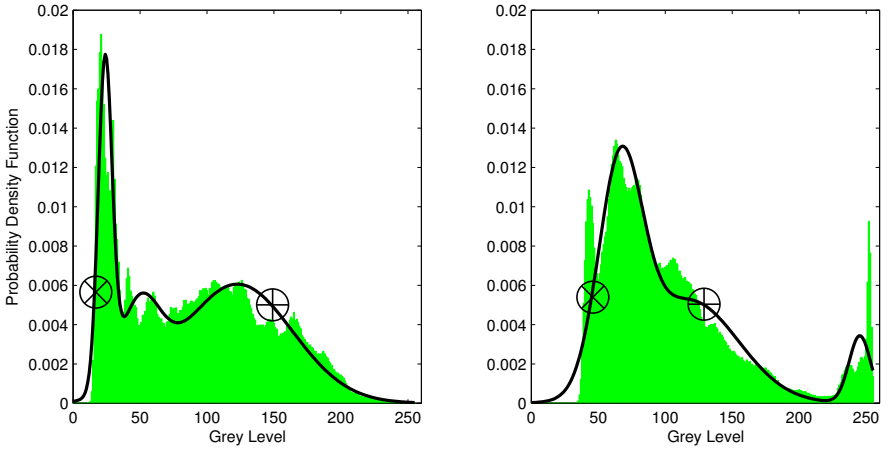


FIGURE 8.2: Histograms (in green) of two images with the estimated PDF (in black) overlaid on top. The marks correspond to the threshold identified for the removal of over/under exposed areas of the image.

reflectance artefacts at different locations, as noted by the manual notation on the images showing NFL and reflectance artefacts enclosed in green and yellow respectively. We also note that the illumination condition is suboptimal in various images.

8.3 Image Processing Method

Our quality enhancer algorithm has four phases: first the images are preprocessed to extract only the information that is likely to be artefact-free; the preprocessed images are initially registered rigidly with a feature based method and then registration is refined with a non-rigid technique that compensate the camera/eye distortions; the areas largely over/under exposed are detected and labelled; finally, the stack of images is statistically analysed at histogram level to extract pixels of interest and the images are combined together. The method to pre-process the image is exactly the one presented in Section 7.3.1, where a description can be found.

8.3.1 Rigid/Non-Rigid Registration

The non rigid registration process starts with the image alignment method (i.e. rigid registration) described in Section 7.3.2. This registration is not able to perfectly match the different views of the fundus because of a slight barrel distortion due to the camera lens and the *lens* inside the eye. The classic solution to this distortion problem is the use of calibration targets to estimate the lens effects, but this is not an option in this application (because it would require placing the calibration targets inside the eye). Therefore,

we further register the images by employing a non-rigid technique that exploits the vessel structure itself to implicitly correct the distortion. In this second phase we leveraged the implementation of the non-rigid registration technique provided by the open source package Elastix (Klein et al., 2010). We employed Mutual Information (Mattes et al., 2003) as similarity metric and the B-splines defined the deformation allowed between the images (as suggested by Rueckert et al. (1999)) with the control points set every 16 pixels in a grid pattern.

8.3.2 Over/Under Exposure Detection

At this stage, we detect the areas that are over or under exposed. This is required in order to avoid the use of these pixels in the generation of the final image in the next phase. First, the VFOV of interest is detected as described in Section 3.2.1.1 (only the pixels inside the VFOV are taken into consideration). For such pixels, the histogram is computed and converted to a probability density function (PDF) by applying Expectation Maximisation (EM) with a Gaussian Mixture model composed of three Gaussians Bishop (2007). This number of Gaussian is empirically chosen in order to roughly model the histogram distribution without the narrow peaks. The EM process is initialised with a simple K-means clustering. Fig. 8.2 shows two examples of the PDFs overlaid on the original histograms. The PDFs approximate the histogram in a way to exclude the thin spikes at the beginning and at the end which are a common symptom of over/under exposure. These areas are detected by finding the first two corresponding histogram bins on the PDF whose probability are above th, starting from the left and the right. Fig. 8.2 shows the boundary detected with an $\otimes$ for the under exposed areas and a $\oplus$ for the over exposed areas. In our experiments th was set to 0.05 through examination of a set of images not used for the final experiment. Binary masks I_m^i containing all the areas below $\otimes$ and above $\oplus$ are generated for each RGB channel of each image. A morphological opening and closing operations are run on all I_m^i to remove the noise and close all the potential holes.

8.3.3 Analysis

At this point we need to combine all the registered images by keeping all the redundant information and discarding all the areas with sudden changes which are likely to contain artefacts. This does not have the risk of removing exudates because of the different reflectivity as shown in Fig. 8.3. In order to facilitate a comparison between the views and to improve the final result we perform a colour normalisation with the

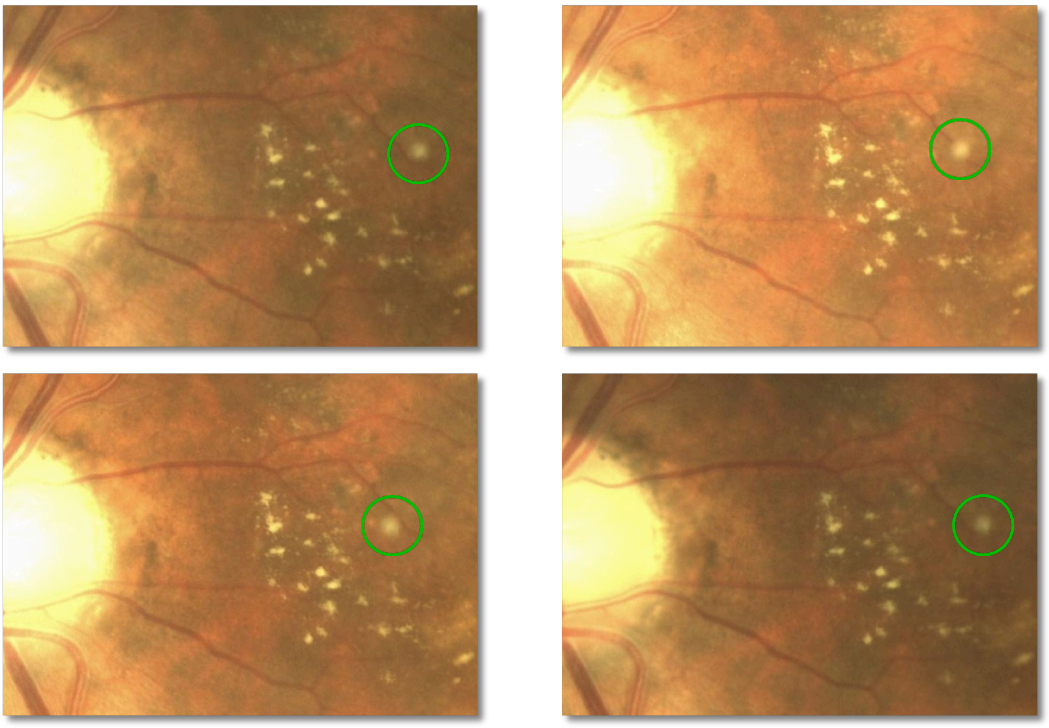


FIGURE 8.3: Details of four fundus images of the same retina captured from different points of view after rigid and non-rigid registration. The NFL reflections are highlighted in green (in this case it might be due to a micro infarction of the NFL, i.e. a cotton wool spot). Note how the position of the exudates remains fixed while the NFL reflection changes its apparent location on the image.

method by Cree et al. (2005), which was described in Section 6.4.2. A detail of an image after this normalisation is shown in Fig. 8.4(b). The images are combined together by the means of a median filtering operation executed through all the images. We stack the images one on top of the other and generate a vector for each pixel as follow $[I(x,y)^0, I(x,y)^1, I(x,y)^2, ...I(x,y)^{n-1}]$ (where n is the number of images, $I(x,y)^i$ a pixel in the image i at the coordinates x,y) and the median of each vector is stored in the final output. Whenever a pixel of I is labelled as being over/under exposed in I_m, it is not employed in the creation of the vector.

8.4 Results

The technique described is applied on a set of 7 fundus images (Fig. 8.5) with a resolution of 1098×979. Fig. 8.4(a) shows the final output obtained. Notice how the vast majority of reflectance artefacts have been removed without affecting the small details. A comparison between 8.4(b) and 8.4(c) makes apparent how the NFL artefact has been

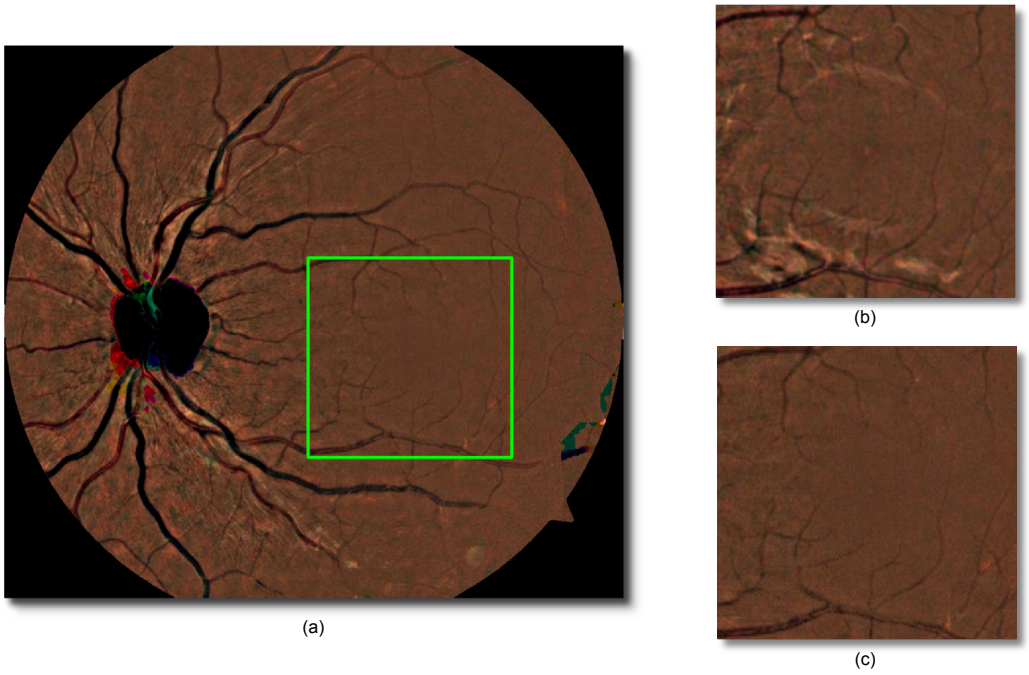


FIGURE 8.4: (a) High quality image generated by our technique employing all the images of Fig. 8.5. (b) Detail of one of the original image (after the colour equalisation described in the analysis stage of Sec. 8.3). (c) Details of the high quality image.

removed without any significant sacrifice in the visibility of even the small blood vessels. At the edge of the image and of the optic disc (the black area at the centre left) some misrepresentations of the colours are visible. This is due to lack of enough areas that do not suffer from over/under exposure for some colour channel, however this does not have any effect in the diagnosis which is mainly concentrated in the macula area (roughly the area highlighted in Fig. 8.4(a)).

Table 8.1 shows the ELVD quality for the initial images compared to the generated image. The output of our technique has a clearly higher score even if the ELVD does not take this type of artefacts in consideration explicitly, but it rather provides a general image quality estimation. We are able to analyse all the images in ∼10 minutes on a

Image 1:	0.564	Image 5:	0.640
Image 2:	0.736	Image 6:	0.538
Image 3:	0.719	Image 7:	0.464
Image 4:	0.411		
Output Image:	0.879		

TABLE 8.1: ELVD QA before and after the automatic quality enhancement

2.2GHz machine with 4GB of memory using an unoptimised Matlab algorithm implementation.

8.5 Discussion

We have presented the preliminary results of a novel technique that aims to remove reflectance artefacts and to increase the overall quality of fundus imaging by an unorthodox multiple imaging approach, that is immediately applicable with all the modern digital fundus cameras. This technique could boost the results of automated diagnosis technique with practically no changes in the algorithms. Obviously, more tests are required to prove the technique in a clinical setting and to quantify the advantages on automatic lesion segmentation algorithms with particular attention to exudates, which are lesion that most resemble NFL reflections. One drawback of eliminating the NFL is that Cotton Wool Spots are located in this layer, hence they are removed too, however thy are the only lesions that could be removed and they are not essential for a prompt diagnosis of DR or DME. Also, the great majority of ideas presented in this chapter can be easily transferred to other imaging domains where reflectance artefacts are a problem and the object imaged can be rotated with respect to the camera.

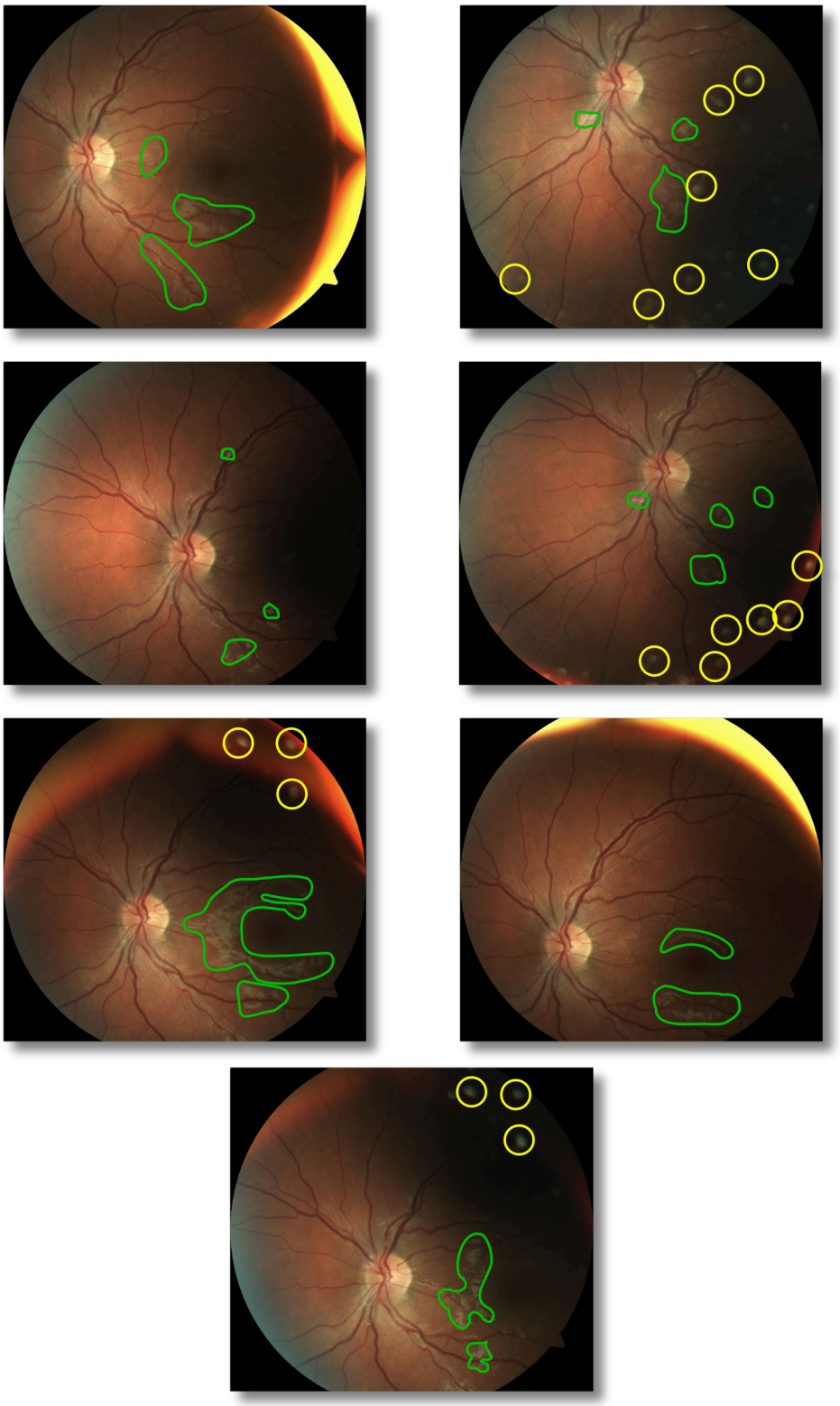


FIGURE 8.5: Images of the same patient acquired by changing the fixation point. In green (or dark grey) the most obvious NFL reflectance; in yellow (or bright grey) the other reflectance artefacts due to the dirt on the lens are highlighted.

Conclusions and Further Work

This work presented a summary of the state of the art algorithms for the automatic analysis of fundus images, with particular attention to the aspects related to diabetic retinopathy and retina morphology. An introduction to the required medical terms and concepts was given, together with the common protocols to manually diagnose and grade diabetic retinopathy with fundus images. Different retina imaging techniques were described and a rationale for the choice fundus images was given.

An automatic quality assessment algorithm able to generate a quality score in a short amount of time was introduced. Its performances were evaluated with a small dataset, but more importantly in a live screening environment of a teleophtalmology network and, according to the feedback received, it proved to be a useful addition.

Two new lesion segmentation algorithms were developed, one to target exudates, one microaneurysms. Both of them have been been compared with similar methods and proved to be have competitive performance. As far as the computation time is concerned, they are able to estimate the lesions in a short amount of time. Additionally, they can be trained very easily: the exudate detector due to its rule-based nature, and the MA detector because it can be trained in a few minutes with a graphical user interface. Additionally, the Radon based method to the detect microaneurysms has the immediate potential of being applied to other medical image processing tasks, where round structures need to be identified in a noisy environment.

The output of the microaneurysms localization algorithm was used to build a classifier for diabetic retinopathy. Its performance has been evaluated with multiple datasets and cross-training datasets. Only the HEI-MED dataset was used for developing purposes all the others were only employed in the final testing phase. In addition, an automatic diabetic macular edema detector based on exudates and other features was developed. Its performance are comparable to the ones of retina experts with respect to sensitivity.

A ray traced model of the fundus camera/eye/retina was built in order to test the feasibility of an algorithm for the detection of edema even before the appearance of exudates. We showed how the algorithm presented can achieve this with multiple views of the retina. Furthermore, an algorithm capable of removing bright artefacts after image acquisition was presented. The result was again achieved by capturing multiple fundus images of the patient that stared at different locations.

Overall, this work has presented several advancements in retina image processing geared toward improving the state-of-the-art in automated screening, from the acquisition phase, through quality assessment/enhancements, lesion detection and disease classification. Hence, there are various directions that can be pursued for further studies. First and foremost, a larger clinical evaluation of the complete single image screening process (from quality estimation to diagnosis) is the first step for the acceptance of the medical community of such automatic screening system. This will allow to better identify the system deficiencies and to propose solutions, possibly by combining existing approaches if not secreted because of commercial reasons. Next, new datasets of multi-view retinas needs to be acquired. One will be employed to quantify the improvements in the automatic diagnosis when the quality enhancement algorithm is applied; the other will be used to further evaluate the swollen macula detector algorithm and to find new image processing techniques to limit as much as possible the artefacts at the edges of the reconstructed naive-height-maps. Also, we foresee new application domains for the Radon-based microaneurysms detector. In fact, there are many medical and industrial imaging problems where small round structures needs to be identified in a noisy background. An example is the automatic localization of nodules in lungs radiographies.

Glossary

Age-related Macular Degeneration degenerative condition affecting the macular or central area of the retina characterized by the presence of drusen.

AMD see Age-related Macular Degeneration.

Bruch's Membrane the innermost layer of the choroid. It provides oxygen and nourishment to the outer layers of the retina.

Choroid vascular layer of the eye.

Cotton Wool Spots lesions appearing as puffy white patches on the retina. They are caused by damage to nerve fibers.

CWS See Cotton Wool Spots.

Diabetes Mellitus chronic, systemic, life-threatening disease characterized by disordered metabolism and abnormally high blood sugar (hyperglycaemia) resulting from low levels of the hormone insulin with or without abnormal resistance to insulin's effects.

Diabetic Macular Edema condition that occurs when fluid and protein deposits collect on or under the macula and causes it to thicken and swell.

Diabetic Retinopathy retinopathy (i.e. damage to the retina) caused by complications of diabetes mellitus, which can eventually lead to blindness.

DM see Diabetes Mellitus.

DME see Diabetic Macular Edema.

DR see Diabetic Retinopathy.

Drusen tiny yellow or white accumulations of extracellular material that build up in Bruch's Membrane of the eye.

Field of View image area that can be acquired by a camera.

FOV see Field of View.

Fovea located in the center of the macula region of the retina it is responsible for sharp central vision because of the largest concentration of cones.

Fundus Camera a specialized low power microscope with an attached camera designed to photograph the interior surface of the eye.

Ganglion Cell type of neuron located near the inner surface (the ganglion cell layer) of the retina.

Glaucoma conditions that affects the retina by damaging ganglion cells and their axons.

Hard Exudates (or exudates) lipid break-down products (appearing as yellow spots in a fundus image) that are left behind after localized edema resolves.

HE see Hard Exudates.

Hemorrhage medical term for bleeding.

HMA see Hemorrhage.

Infarction Tissue death (necrosis) that is caused by a local lack of oxygen due to obstruction of the tissue's blood supply.

MA see Microaneusym.

Macula oval-shaped highly pigmented yellow spot near the center of the retina of the eye.

Microaneurysm localized, blood-filled balloon-like bulge in the wall of a blood vessel.

Microinfarction a very small infarct due to obstruction of circulation in capillaries, arterioles, or small arteries.

Neovascularization vascular abnormality due to the lack of oxigen to an area of the retina.

Nerve Fibre Layer retina layer is formed by the expansion of the fibers of the optic nerve.

NFL see Nerve Fibre Layer.

Optic Disc area representing the beginning of the optic nerve and the entry point for the major blood vessels that supply the retina.

Optic Nerve nerve that transmits visual information from the retina to the brain.

Preretinal Hemorrhages Hemorrhage occuring between vitreous humour and retina.

PRH see Preretinal Hemorrhages.

VH see Vitreous Hemorrhage.

Vitreous Hemorrhage Hemorrhage occurring in the vitreous humour.

Vitreous Humour the clear gel that fills the space between the lens and the retina of the eyeball.

Publications

Journals

- L Giancardo, F Meriaudeau, T P Karnowski, Y Li, S Gaarg, K W Tobin and E Chaum, "Exudate-based Diabetic Macular Edema Detection in Fundus Images Using Publicly Available Datasets", *Medical Image Analysis (in printing)*.
- L Giancardo, F Meriaudeau, T P Karnowski, K W Tobin, P Favaro, E Grisan, A Ruggeri and E Chaum, "Texture-less Macula Swelling Detection with Multiple Retinal Fundus Images", *IEEE Transactions on Biomedical Engineering Letters, March 2011*.
- Y Li, T P Karnowski, K W Tobin, L Giancardo, S Morris, E Sparrow, S Garg, K Fox and E Chaum, "A HIPAA-Compliant Ocular Telehealth Network for the Remote Diagnosis and Management of Diabetic Retinopathy", *Telemedicine and e-Health (in printing)*.

Book Chapters

- L Giancardo, F Meriaudeau, T P Karnowski, E Chaum and K W Tobin, "Quality Assessment of Retinal Fundus Images using ELVD". *IN-TECH, New Developments in Biomedical Engineering, pp. 201–224, 2010*.
- Thomas P. Karnowski, Yaqin Li, Luca Giancardo, Deniz Aykac, Seema Garg, Michael D. Abramoff, Matthew T. Tennant, Kenneth W. Tobin and Edward Chaum "Automating the Diagnosis, Stratification, and Management of Diabetic Retinopathy

Using Content-Based Image Retrieval in an Ocular Telehealth Network". *Computational Analysis of the Human Eye with Applications, World Scientific Publications, 2010*.

Conference Proceedings

- L Giancardo, F Meriaudeau, T P Karnowski, Y Li, K W Tobin and E Chaum, "Microaneurysm Detection with Radon Transform Based Classification on Retina Fundus Images". *IEEE EMBC, 2011*.
- L Giancardo, F Meriaudeau, T P Karnowski, Y Li, K W Tobin and E Chaum, "Automatic Quality Enhancement And Nerve Fibre Layer Artefacts Removal In Retina Fundus Images By Off Axis Imaging ". *IEEE International Conference on Image Processing, 2011*.
- L Giancardo, F Meriaudeau, T P Karnowski, Y Li, K W Tobin and E Chaum, "Automatic Retina Exudates Segmentation without a Manually Labelled Training Set". *Proceedings of International Symposium on Biomedical Imaging, 2011*.
- L Giancardo, F Meriaudeau, T P Karnowski, E Chaum, K W Tobin and Y Li, "Microaneurysms Detection with the Radon Cliff Operator in Retinal Fundus Images". *Proceedings of SPIE on Medical Imaging, 2010*.
- L Giancardo, F Meriaudeau, T P Karnowski, E Chaum, K W Tobin and Y Li, "Bright Retinal Lesions Detection using Color Fundus Images Containing Reflective Features ". *Proceedings of World Congress, 2009, Munich, Germany*.
- L Giancardo, F Meriaudeau, T P Karnowski, E Chaum, K W Tobin and M Abramoff, "Elliptical Local Vessel Density: a Fast and Robust Quality Metric for Retinal Images". *IEEE EMBC, 2008, Vancouver, Canada*.
- T P Karnowski, D Aykac, L Giancardo, T L Nichols, Y Li, K W Tobin and E Chaum, "Automatic Detection of Retina Disease: Robustness to Image Quality and Localization of Anatomy Structure". IEEE EMBC, 2011.
- K Devisetti, T P Karnowski, L Giancardo, Y Li and E Chaum, "Geographic Atrophy Segmentation in Infra-Red and Auto Fluorescent Retina Images using Supervised Learning". *IEEE EMBC, 2011*.
- H Santos-Villalobos, T P Karnowski, D Aykac, L Giancardo, Y Li, T L Nichols, K W Tobin and E Chaum, "Statistical Characterization and Segmentation of Drusen in Fundus Images". *IEEE EMBC, 2011*.

- A Boone, T P Karnowski, E Chaum, L Giancardo, Y Li and K W Tobin , "Image Processing and Hierarchical Temporal Memories for Automated Retina Analysis". *Proceedings of BSEC, Oak Ridge, TN*.

- T P Karnowski, Deniz Aykac, E Chaum, L Giancardo, Y Li and K W Tobin, "Evaluating the Accuracy of Optic Nerve Detection in Retina Imaging Using Complementary Methods ". *Proceedings of World Congress 2009, Munich, Germany*.

- T P Karnowski, Deniz Aykac, E Chaum, L Giancardo, Y Li and K W Tobin, "Practical Considerations for Optic Nerve Location in Telemedicine ". *IEEE EMBC 2009*.

- K W Tobin, M Abramoff, E Chaum, L Giancardo, V Govindasamy, T P Karnowski, M Tennant and S Swainson, "Using a Patient Image Archive to Diagnose Retinopathy ". *IEEE EMBC, 2008*.

Posters

- L Giancardo, F Meriaudeau, T P Karnowski, Y Li , S Garg, K Fox, K W Tobin and E Chaum, "A Comparison between a new Automated Macular Edema Screening System and Retina Experts Diagnosis". *ARVO Annual Meeting, Fort Lauderdale, FL, 2011*.

- L Giancardo, F Meriaudeau, T P Karnowski, Y Li , K W Tobin and E Chaum, "An Algorithm for Automated Diagnosis of Clinically Significant Macular Edema in a Teleophthalmology Network". *ARVO Annual Meeting, Fort Lauderdale, FL, 2010*.

- L Giancardo, F Meriaudeau, T P Karnowski, E Chaum, K W Tobin and Y Li, "Exudates Detection with a Hybrid Background Analysis in Retinal Fundus Images ". *Memphis BioImaging Symposium, 2009*.

- Y Li, T P Karnowski, K W Tobin, L Giancardo and E Chaum, "Diabetic Retinopathy Management in the Mid-South Using a Semi-Automated Teleophthalmology Network: RISA-NET ". *American Telemedicine Association, 2010*.

- Y Li, T P Karnowski, K W Tobin, L Giancardo and E Chaum, "A Network Infrastructure for Automated Diagnosis of Diabetic Retinopathy ". *American Telemedicine Association, 2009*.

References

Abramoff, M. D., Garvin, M. K., and Sonka, M. (2010a). Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, 3:169–208. (Cited in section 2.1.1, 2.1.2, and 2.1.4.4.)

Abramoff, M. D., Reinhardt, J. M., Russell, S. R., Folk, J. C., Mahajan, V. B., Niemeijer, M., and Quellec, G. (2010b). Automated early detection of diabetic retinopathy. *Ophthalmology*, 117(6):1147–1154. (Cited in section 2.4.2.)

Abramoff, M. D. and Suttorp-Schulten, M. S. A. (2005). Web-based screening for diabetic retinopathy in a primary care population: the eyecheck project. *Telemed J E Health*, 11(6):668–674. (Cited in section 2.4.2.)

Agurto, C., Murray, V., Barriga, E., Murillo, S., Pattichis, M., Davis, H., Russell, S., Abramoff, M., and Soliz, P. (2010). Multiscale am-fm methods for diabetic retinopathy lesion detection. *IEEE Transaction on Medical Imaging*, 29(2):502–512. (Cited in section 6.7 and 6.8.)

Aldington, S. J., Kohner, E. M., Meuer, S., Klein, R., and Sjølie, A. K. (1995). Methodology for retinal photography and assessment of diabetic retinopathy: the eurodiab iddm complications study. *Diabetologia*, 38(4):437–444. (Cited in section 2.4.1.)

Antal, B., Lazar, I., and Hajdu, A. (2010). An ensemble-based system for microaneurysm detection. Technical report, University of Debrecen, Hungary. (Cited in section 2.3.1, 2.3.1, 5.1, and 5.5.)

Aquino, A., Gegundez-Arias, M. E., and Marin, D. (2010). Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Transactions on Medical Imaging*, 29(11):1860–1869. (Cited in section 2.2.4.)

ARIA (2006). Retinal image archive. Online [http://www.eyecharity.com]. (Cited in section 2.2.1.)

Baker, M. L., Hand, P. J., Wang, J. J., and Wong, T. Y. (2008). Retinal signs and stroke: revisiting the link between the eye and brain. *Stroke*, 39(4):1371–1379. (Cited in section 2.1.4.1.)

Baudoin, C. E., Lay, B. J., and Klein, J. C. (1984). Automatic detection of microaneurysms in diabetic fluorescein angiography. *Rev Epidemiol Sante Publique*, 32(3-4):254–261. (Cited in section 2.3.1.)

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Surf: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359. (Cited in section 7.3.2.)

Bishop, C. (2007). *Pattern Recognition and Machine Learning*. Springer, 2nd edition. (Cited in section 8.3.2.)

Bouguet, J. (2000). Pyramidal implementation of the lucas kanade feature tracker description of the algorithm. Technical report, Intel Corporation, Microprocessor Research. (Cited in section 7.3.3.)

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32. (Cited in section 6.4.5.)

Cassin, B. and Solomon, S. (1990). *Dictionary of Eye Terminology*. Gainsville, Florida: Triad Publishing Company. (Cited in section 2.1.1.)

Cattin, P. C., Bay, H., Gool, L. V., and Szekely, G. (2006). Retina mosaicing using local features. *Med Image Comput Comput Assist Interv*, 9(Pt 2):185–192. (Cited in section 7.3.2.)

Centers for Disease Control and Prevention (2011). National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the united states. Technical report, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. (Cited in section 1, 2.1.4, and 2.4.2.)

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. (Cited in section 5.2.4 and 6.4.5.)

Chanwimaluang, T., Fan, G., Yen, G. G., and Fransen, S. R. (2009). 3-d retinal curvature estimation. *IEEE Trans Inf Technol Biomed*, 13(6):997–1005. (Cited in section 7.1.)

Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., and Goldbaum, M. (1989). Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transaction on Medical Imaging*, 8(3):263–269. (Cited in section 2.2.3.)

Chernecky, C. and Berger, B. (2004). *Laboratory Tests and Diagnostic Procedures*. Philadelphia: Saunders, 4th edition. (Cited in section 2.1.2.)

Choe, T. E., Cohen, I., and Medioni, G. (2006). 3-d shape reconstruction of retinal fundus. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2277–2284. (Cited in section 7.1.)

Cinsdikici, M. G. and Aydin, D. (2009). Detection of blood vessels in ophthalmoscope images using mf/ant (matched filter/ant colony) algorithm. *Comput. Methods Programs Biomed.*, 96(2):85–95. (Cited in section 2.2.3.)

Corona, E., Mitra, S., Wilson, M., Krile, T., Kwon, Y. H., and Soliz, P. (2002). Digital stereo image analyzer for generating automated 3-d measures of optic disc deformation in glaucoma. *IEEE Transactions on Medical Imaging*, 21(10):1244–1253. (Cited in section 7.1.)

Cree, M. J., Gamble, E., and Cornforth, D. (2005). Colour normalisation to reduce interpatient and intra-patient variability in microaneurysm detection in colour retinal images. In *WDIC2005 ARPS Workshop on Digital Image Computing*. (Cited in section 2.2.2, 5.2.2, 6.4.2, and 8.3.3.)

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley-Interscience. (Cited in section 4.3.2.1 and 6.4.5.)

Durrant-Whyte, H. and Bailey, T. (2006). Simultaneous localisation and mapping (slam): Part i the essential algorithms. *Robotics and Automation Magazine*, 13(2):99–110. (Cited in section 7.1.)

Early Treatment Diabetic Retinopathy Study Research Group (1991). Grading diabetic retinopathy from stereoscopic color fundus photographs–an extension of the modified airlie house classification. etdrs report number 10. early treatment diabetic retinopathy study research group. *Ophthalmology*, 98(5 Suppl):786–806. (Cited in section 2.4.1 and 2.5.)

Farsiu, S., Elad, M., and Milanfar, P. (2006). Multiframe demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, 15(1):141–159. (Cited in section 8.1.)

Fawcett, T. (2004). Roc graphs : Notes and practical considerations for researchers. Technical report, HP Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304, USA. (Cited in section 2.2.1.)

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874. (Cited in section 2.2.1.)

Feng, P., Pan, Y., Wei, B., Jin, W., and Mi, D. (2007). Enhancing retinal image by the contourlet transform. *Pattern Recognition Letters*, 28:516–522. (Cited in section 2.2.2.)

Fleming, A. D., Goatman, K. A., and Philip, S. (2010). The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy. *British Journal of Ophthalmology*. (Cited in section 2.3.2.)

Fleming, A. D., Goatman, K. A., Philip, S., Olson, J. A., and Sharp, P. F. (2007). Automatic detection of retinal anatomy to assist diabetic retinopathy screening. *Physics in Medicine and Biology*, 52(2):331–345. (Cited in section 2.2.4.)

Fleming, A. D., Philip, S., Goatman, K. A., Olson, J. A., and Sharp, P. F. (2006a). Automated assessment of diabetic retinal image quality based on clarity and field definition. *Invest Ophthalmol Vis Sci*, 47(3):1120–1125. (Cited in section 2.2.5, 3.1, and 3.2.2.1.)

Fleming, A. D., Philip, S., Goatman, K. A., Olson, J. A., and Sharp, P. F. (2006b). Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE Transactions on Medical Imaging*, 25(9):1223–1232. (Cited in section 4.4.1.)

Foracchia, M., Grisan, E., and Ruggeri, A. (2004). Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transaction on Medical Imaging*, 23(10):1189–1195. (Cited in section 2.2.4.)

Foracchia, M., Grisan, E., and Ruggeri, A. (2005). Luminosity and contrast normalization in retinal images. *Medical Image Analysis*, 9(3):179–190. (Cited in section (document), 2.2.2, 4.3.2.1, 4.2, and 4.4.1.)

Frame, J. A., Undrill, P. E., Cree, M. J., Olson, J. A., McHardy, K. C., Sharp, P. F., and Forrester, J. V. (1998). A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms. *Computers in Biology and Medicine*, 28:225–238. (Cited in section 2.3.1.)

Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906. (Cited in section 6.3.)

Garcia, M., Sanchez, C. I., Lopez, M. I., Abasolo, D., and Hornero, R. (2009). Neural network based detection of hard exudates in retinal images. *Comput Methods Programs Biomed*, 93(1):9–19. (Cited in section 2.3.2.)

Gardner, G. G., Keating, D., Williamson, T. H., and Elliott, A. T. (1996). Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *Br J Ophthalmol*, 80(11):940–944. (Cited in section 2.3.2.)

Giancardo, L., Abramoff, M., Chaum, E., Karnowski, T., Meriaudeau, F., and Tobin, K. (2008). Elliptical local vessel density: a fast and robust quality metric for retinal images. In *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2008*. (Cited in section 3.)

Giancardo, L., Meriaudeau, F., Karnowski, T., Li, Y., Tobin Jr, K., and Chaum, E. (2011a). Automatic quality enhancement and nerve fibre layer artefacts removal in retina fundus images by off axis imaging. In *IEEE International Conference on Image Processing*. (Cited in section 8.)

Giancardo, L., Meriaudeau, F., Karnowski, T., Li, Y., Tobin Jr, K., and Chaum, E. (2011b). Microaneurysm detection with radon transform-based classification on retina images. In *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2011 (submitted)*. (Cited in section 5.)

Giancardo, L., Meriaudeau, F., Karnowski, T., Tobin Jr, K., Li, Y., and Chaum, E. (2010a). Microaneurysms detection with the radon cliff operator in retinal fundus images. In *Proceedings of SPIE*, volume 7623, page 76230U. (Cited in section 2.3.1, 2.3.1, 5, and 5.1.)

Giancardo, L., Meriaudeau, F., Karnowski, T. P., Chaum, E., and Tobin, K. (2010b). *Quality Assessment of Retinal Fundus Images using ELVD*, chapter New Developments in Biomedical Engineering, pages 201–224. IN-TECH. (Cited in section 2.2.5 and 3.)

Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Garg, S., Tobin Jr, K., and Chaum, E. (2011c). Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis (in printing)*. (Cited in section 2.2.1, 2.3.2, 4, and 6.)

Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Tobin Jr, K., and Chaum, E. (2011d). Automatic retina exudates segmentation without a manually labelled training set. In *Proc. of IEEE International Symposium on Biomedical Imaging*. (Cited in section 2.2.2, 2.3.2, and 4.)

Giancardo, L., Meriaudeau, F., Karnowski, T. P., Tobin, K. W., Grisan, E., Favaro, P., Ruggeri, A., and Chaum, E. (2011e). Textureless macula swelling detection with multiple retinal fundus images. *IEEE Transactions on Biomedical Engineering*, 58(3):795–799. (Cited in section 7.)

Gonzales, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Prentice-Hall. (Cited in section 2.2.2 and 3.2.1.1.)

Gregson, P. H., Shen, Z., Scott, R. C., and Kozousek, V. (1995). Automated grading of venous beading. *Comput Biomed Res*, 28(4):291–304. (Cited in section 2.1.3.6.)

Grisan, E. (2005). *Automatic Analysis of Retinal Images: Retinopathy Detection and Grading*. PhD thesis, Università di Padova. (Cited in section 2.1.3.6.)

Grisan, E., Foracchia, M., and Ruggeri, A. (2008). A novel method for the automatic grading of retinal vessel tortuosity. *IEEE Transaction on Medical Imaging*, 27(3):310–319. (Cited in section 2.2.3.)

Gwet, K. (2002). Inter-rater reliability: dependency on trait prevalence and marginal homogeneity. *Statistical Methods For Inter-Rater Reliability Assessment Series*, 2:1–9. (Cited in section 2.2.1.)

Halir, R. and Flusser, J. (2000). Numerically stable direct least squares fitting of ellipses. *Department of Software Engineering, Charles University, Czech Republic*. (Cited in section 3.2.1.2.)

Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press. (Cited in section 7.1 and 7.3.2.)

Heneghan, C., Flynn, J., O'Keefe, M., and Cahill, M. (2002). Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Med Image Anal*, 6(4):407–429. (Cited in section 3.2.1.3.)

Hoover, A. and Goldbaum, M. (2003). Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transactions on Medical Imaging*, 22(8):951–958. (Cited in section 2.2.1 and 2.2.4.)

Hoover, A. D., Kouznetsova, V., and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210. (Cited in section 2.2.3.)

Hutchinson, A., McIntosh, A., Peters, J., O'Keeffe, C., Khunti, K., Baker, R., and Booth, A. (2000). Effectiveness of screening and monitoring tests for diabetic retinopathy–a systematic review.. jul;17(7):. *Diabet Med*, 17:495–506. (Cited in section 2.1.2.)

Intel (2007). *Intel Integrated Performance Primitives for the Windows OS on the IA-32 Architecture*, 318254-001us edition. (Cited in section 3.3.4.)

Jain, A. K., Duin, R. P. W., and Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37. (Cited in section 6.4.5.)

Jiang, X. and Mojon, D. (2003). Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):131–137. (Cited in section 2.2.3.)

Joint Committee for Guides in Metrology (2004). International vocabulary of basic and general terms in metrology (vim). Technical report. (Cited in section 2.2.1.)

Jonas, J., Schneider, U., and Naumann, G. (1992). Count and density of human retinal photoreceptors. *Graefe's Arch Clin Exp Ophthalmol*, pages 230:505–510. (Cited in section 2.1.1.)

Joussen, A., Gardner, T., and Kirchhof, B. (2007). *Retinal vascular disease*. Springer Verlag. (Cited in section 2.1.4.1, 2.1.4.2, and 2.4.1.)

Karnowski, T. P., Aykac, D., Chaum, E., Giancardo, L., Li, Y., Tobin, K. W., and Abramoff, M. D. (2009). Practical considerations for optic nerve location in telemedicine. In *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2009*, pages 6205–6209. (Cited in section 2.2.4 and 3.4.)

Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., L, L., Sorri, I., Uusitalo, H., Pietila, J., Kalviainen, H., and Uusitalo, H. (2007). the diaretdb1 diabetic retinopathy database and evaluation protocol. In *Proceedings of British Machine Vision Conference*. (Cited in section 2.2.1.)

Kingston, A. and Autrusseau, F. (2008). Lossless image compression via predictive coding of discrete radon projections. *Signal Processing: Image Communication*. (Cited in section 5.1.)

Kirsch, R. A. (1971). Computer determination of the constituent structure of biological images. *Computers and Biomedical Research*, 4(3):315–328. (Cited in section 4.3.2.2 and 4.4.2.)

Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. (2010). elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205. (Cited in section 8.3.1.)

Lalonde, M., Gagnon, L., and Boucher, M. (2001). Automatic visual quality assessment in optical fundus images. *Proceedings of Vision Interface*. (Cited in section 2.2.5, 3.1, and 3.2.3.)

Lazar, I. and Hajdu, A. (2011). Microaneurisms detection in retinal images using a rotating cross-section based model. In *Proc. of IEEE International Symposium on Biomedical Imaging*. (Cited in section 2.3.1, 2.3.1, and 5.3.)

Leavers, V. F. (1992). Shape detection in computer vision using the hough transform. *Springer-Verlag*. (Cited in section 3.2.1.2.)

Lee, S. and Wang, Y. (1999). Automatic retinal image quality assessment and enhancement. In *Proceedings of SPIE Image Processing.*, pages 1581–1590. (Cited in section 2.2.5, 3.1, 3.2.2.2, and 3.2.3.)

Li, H. and Chutatape, O. (2004). Automated feature extraction in color retinal images by a model based approach. *IEEE Transactions on Biomedical Engineering*, 51(2):246–254. (Cited in section 2.3.2.)

Li, H. K., Hubbard, L. D., Danis, R. P., Esquivel, A., Florez-Arango, J. F., and Krupinski, E. A. (2010). Monoscopic vs. stereoscopic retinal photography for grading diabetic retinopathy severity. *Invest Ophthalmol Vis Sci*. (Cited in section 7.1.)

Li, Y., Karnowski, T., Tobin, K., Giancardo, L., Morris, S., Sparrow, S., Garg, S., Fox, K., and Chaum, E. (2011). A hipaa-compliant ocular telehealth network for the remote diagnosis and management of diabetic retinopathy. *Telemedicine and eHealth (in press)*. (Cited in section 2.4.2, 3.4, and 4.2.)

Lin, D., Blumenkranz, M., Brothers, R., and Grosvenor, D. (2002). The sensitivity and specificity of singlefield nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *American Journal of Ophthalmology*, 134:261–3. (Cited in section 2.1.2.)

Lipson, B. K. and Yannuzzi, L. A. (1989). Complications of intravenous fluorescein injections. *Int Ophthalmol Clin*, 29(3):200–205. (Cited in section 2.3.1.)

Liu, D., Wood, N. B., Xu, X. Y., Witt, N., Hughes, A. D., and SAMcG, T. (2008). *3D Reconstruction of the Retinal Arterial Tree Using Subject-Specific Fundus Images*, chapter Advances in Computational Vision and Medical Image Processing, pages 187–201. SpringerLink. (Cited in section 7.1.)

Liu, G., Lin, Z., and Yu, Y. (2009). Radon representation-based feature descriptor for texture classification. *IEEE Trans on Image Proc*, 18:921–928. (Cited in section 5.1.)

Luzio, S., Hatcher, S., Zahlmann, G., Mazik, L., Morgan, M., and Liesenfeld, B. (2004). Feasibility of using the tosca telescreening procedures for diabetic retinopathy. *Diabet Med.*, 21:1121–8. (Cited in section 2.1.2.)

Mallat, S. (1999). *A wavelet tour of signal processing*. Academic press. (Cited in section 4.4.2.2.)

Marin, D., Aquino, A., Gegundez-Arias, M. E., and Bravo, J. M. (2011). A new supervised method for blood vessel segmentation in retinal images by using gray-level and

moment invariants-based features. *IEEE Transaction on Medical Imaging*, 30(1):146–158. (Cited in section 2.2.3.)

Martinello, M., Favaro, P., Muyo Nieto, G. D., Harvey, A. R., Grisan, E., Scarpa, F., and Ruggeri, A. (2007). 3-d retinal surface inference: Stereo or monocular fundus camera? In *Proc. of EMBS*, pages 896–899. (Cited in section 7.1.)

Martinez-Perez, M. E., Hughes, A. D., Thom, S. A., Bharath, A. A., and Parker, K. H. (2007). Segmentation of blood vessels from red-free and fluorescein retinal images. *Med Image Anal*, 11(1):47–61. (Cited in section 2.2.3.)

Masland, R. H. (2001). The fundamental plan of the retina. *Nat Neurosci*, 4(9):877–886. (Cited in section 2.1.1.)

Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). Pet-ct image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128. (Cited in section 8.3.1.)

Melville, A., Richardson, R., McIntosh, A., O'Keeffe, C., Mason, J., Peters, J., and Hutchinson, A. (2000). Complications of diabetes: screening for retinopathy and management of foot ulcers. *Qual Health Care*, 9(2):137–141. (Cited in section 3.)

Mendonca, A. M. and Campilho, A. (2006). Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging*, 25(9):1200–1213. (Cited in section 2.2.3.)

Messidor (2010). Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. Online [http://messidor.crihan.fr]. (Cited in section 2.2.1.)

Metz, C. E. (1986). Roc methodology in radiologic imaging. *Invest Radiol*, 21(9):720–733. (Cited in section 2.2.1.)

Mizutani, A., Muramatsua, C., Hatanakab, Y., Suemoria, S., Haraa, T., and Fujitaa, H. (2009). Automated microaneurysm detection method based on double-ring filter in retinal fundus images. *Conf Proc of SPIE*, 7260. (Cited in section 2.3.1 and 2.3.1.)

Nakagawa, T., Suzuki, T., Hayashi, Y., Mizukusa, Y., Hatanaka, Y., Ishida, K., Hara, T., Fujita, H., and Yamamoto, T. (2008). Quantitative depth analysis of optic nerve head using stereo retinal fundus image pair. *J Biomed Opt*, 13(6):064026. (Cited in section 7.1.)

Niemeijer, M., Abramoff, M. D., and van Ginneken, B. (2006). Image structure clustering for image quality verification of color retina images in diabetic retinopathy

screening. *Med Image Anal*, 10(6):888–898. (Cited in section 2.2.5, 3.1, 3.2.2.2, 3.2.3, 3.3.2, and 3.3.4.)

Niemeijer, M., Abramoff, M. D., and van Ginneken, B. (2009a). Information fusion for diabetic retinopathy cad in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 28(5):775–785. (Cited in section 2.4.2 and 6.8.)

Niemeijer, M., Abràmoff, M. D., and van Ginneken, B. (2009b). Fast detection of the optic disc and fovea in color fundus photographs. *Medical Image Analysis*, 13(6):859–870. (Cited in section 2.2.4.)

Niemeijer, M., Garvin, M., van Ginneken, B., Sonka, M., and Abramoff, M. (2008). Vessel segmentation in 3d spectral oct scans of the retina. In *Proceedings of SPIE*, volume 6914. (Cited in section 2.1.2.)

Niemeijer, M., Niemeijer, M., Abramoff, M. D., and van Ginneken, B. (2007a). Segmentation of the optic disc, macula and vascular arch in fundus photographs. 26(1):116–127. (Cited in section 2.2.4.)

Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., and Abramoff, M. D. (2004). Comparative study of retinal vessel segmentation methods on a new publicly available database. *Proceeding of SPIE Medical Imaging*, 5370:648–657. (Cited in section 2.2.1 and 2.2.3.)

Niemeijer, M., van Ginneken, B., Cree, M. J., Mizutani, A., Quellec, G., Sanchez, C. I., Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., Wu, X., Cazuguel, G., You, J., Mayo, A., Li, Q., Hatanaka, Y., Cochener, B., Roux, C., Karray, F., Garcia, M., Fujita, H., and Abramoff, M. D. (2010). Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 29(1):185–195. (Cited in section 2.2.1, 2.2.1, and 2.3.1.)

Niemeijer, M., Van Ginneken, B., Russell, S. R., Suttorp-Schulten, M. S. A., and Abramoff, M. D. (2007b). Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest Ophthalmol Vis Sci*, 48(5):2260–2267. (Cited in section 2.3.2.)

Niemeijer, M., van Ginneken, B., Staal, J., Suttorp-Schulten, M. S. A., and Abramoff, M. D. (2005). Automatic detection of red lesions in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 24(5):584–592. (Cited in section 2.3.1, 2.3.1, 4.4.1, 5.1, and 5.5.)

Niki, T., Muraoka, K., and Shimizu, K. (1984). Distribution of capillary non-perfusion in early-stage diabetic retinopathy. *Ophthlamology*, 91(12):1431–1439. (Cited in section 2.1.3.2.)

Ojala, T. and Pietikainen, M. (1996). A comparative study of texture measures with classification based on feature distribution. *Pattern Recognition*, 29:51–59. (Cited in section 3.2.2.3.)

Osareh, A., Mirmehdi, M., Thomas, B., and Markham, R. (2003). Automated identification of diabetic retinal exudates in digital colour images. *Br J Ophthalmol*, 87(10):1220–1223. (Cited in section 2.3.2.)

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66. (Cited in section 4.3.1.2.)

Park, S. C., Park, M. K., and Kang, M. G. (2003). Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36. (Cited in section 8.1.)

Parmet, S. (2001). Age-related macular degeneration. *JAMA*, 288(18). (Cited in section 2.1.4.3.)

Patton, N., Aslam, T. M., MacGillivray, T., Deary, I. J., Dhillon, B., Eikelboom, R. H., Yogesan, K., and Constable, I. J. (2006). Retinal image analysis: concepts, applications and potential. *Progress in Retinal and Eye Research*, 25(1):99–127. (Cited in section 3.1.)

Paulus, J., Meier, J., Bock, R., Hornegger, J., and Michelson, G. (2010). Automated quality assessment of retinal fundus photos. *Int J Comput Assist Radiol Surg*, 5(6):557–564. (Cited in section 2.2.5.)

Philip, S., Fleming, A. D., Goatman, K. A., Fonseca, S., McNamee, P., Scotland, G. S., Prescott, G. J., Sharp, P. F., and Olson, J. A. (2007). The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol*, 91(11):1512–1517. (Cited in section 2.4.2.)

Phillips, R., Forrester, J., and Sharp, P. (1993). Automated detection and quantification of retinal exudates. *Graefes Arch Clin Exp Ophthalmol*, 231(2):90–94. (Cited in section 2.3.2.)

Quellec, G., Lamard, M., Josselin, P. M., Cazuguel, G., Cochener, B., and Roux, C. (2008). Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Transaction on Medical Imaging*, 27(9):1230–1241. (Cited in section 2.3.1, 2.3.1, 4.4.2.2, 5.3, and 5.4.)

Quellec, G., Russell, S. R., and Abramoff, M. D. (2011). Optimal filter framework for automated, instantaneous detection of lesions in retinal images. *IEEE Transactions on Medical Imaging*, 30(2):523–533. (Cited in section 5.3.)

Resnick, J. (1985). The radon transforms and some of its applications. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(1):338–339. (Cited in section 5.1.)

Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. P., and Mariotti, S. P. (2004). Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization*, 82:844–851. (Cited in section 2.1.4.4.)

Ricci, E. and Perfetti, R. (2007). Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Transactions on Medical Imaging*, 26(10):1357–1365. (Cited in section 2.2.3.)

Rosin, P. L. (1993). Ellipse fitting by accumulating five-point fits. In *Pattern Recognition Letters*, volume 14, pages 661–699. (Cited in section 3.2.1.2.)

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721. (Cited in section 8.3.1.)

Saine, P. (2006). Fundus photography: What is a fundus camera? *Ophthalmic Photographers' Society*. (Cited in section 2.1.2.)

Sanchez, C. I., Garcia, M., Mayo, A., Lopez, M. I., and Hornero, R. (2009a). Retinal image analysis based on mixture models to detect hard exudates. *Medical Image Analysis*, 13(4):650–658. (Cited in section 2.3.2, 4.1, 4.3, 4.4.1, 4.5, and 4.5.)

Sanchez, C. I., Hornero, R., Mayo, A., and Garcia, M. (2009b). Mixture model-based clustering and logistic regression for automatic detection of microaneurysms in retinal images. *Conf Proc of SPIE*, 7260. (Cited in section 2.3.1 and 2.3.1.)

Singer, D. E., Nathan, D. M., Fogel, H. A., and Schachat, A. P. (1992). Screening for diabetic retinopathy. *Ann Intern Med*, 116(8):660–671. (Cited in section 2.1.4.2.)

Sinthanayothin, C., Boyce, J. F., Williamson, T. H., Cook, H. L., Mensah, E., Lal, S., and Usher, D. (2002). Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Medicine*, 19(2):105–112. (Cited in section 2.3.2.)

Soares, J. V. B., Leandro, J. J. G., Cesar, R. M., Jelinek, H. F., and Cree, M. J. (2006). Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222. (Cited in section 2.2.3.)

Sopharak, A., Uyyanonvara, B., Barman, S., and Williamson, T. H. (2008). Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics*, 32(8):720–727. (Cited in section 2.2.2, 2.3.2, 4.1, 4.3, 4.5, and 4.5.)

Spencer, T., Olson, J. A., McHardy, K. C., Sharp, P. F., and Forrester, J. V. (1996). 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. (Cited in section 2.2.2 and 2.3.1.)

Staal, J., Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., and van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509. (Cited in section 2.2.3.)

Stanton, A. V., Wasan, B., Cerutti, A., Ford, S., Marsh, R., Sever, P. P., Thom, S. A., and Hughes, A. D. (1995). Vascular network changes in the retina with age and hypertension. *Journal of Hypertension*, 13(12 Pt 2):1724–1728. (Cited in section 2.2.3.)

Teng, T., Lefley, M., and Claremont, D. (2002). Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. *Med Biol Eng Comput*, 40(1):2–13. (Cited in section 3.2.1.1.)

ter Haar Romeny (2003). *Front-End Vision and Multi-Scale Image Analysis*. Springer, 1st edition. (Cited in section 2.2.5.)

Tierney, L. M., McPhee, S. J., and Papadakis, M. A. (2002). *Current medical Diagnosis & Treatment. International edition.* New York: Lange Medical Books/McGraw-Hill. (Cited in section 2.1.4.)

Tobin, K. W., Chaum, E., Govindasamy, V. P., and Karnowski, T. P. (2007). Detection of anatomic structures in human retinal imagery. *IEEE Transactions on Medical Imaging*, 26(12):1729–1739. (Cited in section 2.2.4, 6.1, 6.3, and 6.5.)

Tramontan, L., Poletti, E., Fiorin, D., and Ruggeri, A. (2011). A web-based system for the quantitative and reproducible assessment of clinical indexes from the retinal vasculature. *IEEE Transactions on Biomedical Engineering*, 58(3):818–821. (Cited in section 2.2.3.)

Usher, D., Himaga, M., and Dumskyj, M. (2003). Automated assessment of digital fundus image quality using detected vessel area. In *Proceedings of Medical Image Understanding and Analysis*, pages 81–84. British Machine Vision Association (BMVA). (Cited in section 2.2.5, 3.2, 3.2.2.1, and 3.2.3.)

Vijan, S., Hofer, T. P., and Hayward, R. A. (2000). Cost-utility analysis of screening intervals for diabetic retinopathy in patients with type 2 diabetes mellitus. *JAMA*, 283(7):889–896. (Cited in section 3.)

Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Journal of Image Processing*, 2(2):176–201. (Cited in section 3.2.1.3, 4.3.1.1, 4.3.1.2, and 4.4.1.)

Walsh, A. C., Wildey, R., Lara, C., Ouyang, Y., and Sadda, S. R. (2010). Detection of fundus abnormalities using 3d-oct versus mydriatic color fundus imaging. In *ARVO 2010*. (Cited in section 2.1.2.)

Walter, T., Klein, J. K., Massin, P., and Erginay, A. (2002). A contribution of image processing to the diagnosis of diabetic retinopathy–detection of exudates in color fundus images of the human retina. *IEEE Transactions on Medical Imaging*, 21(10):1236–1243. (Cited in section 2.3.2.)

Wang, Y., Tan, W., and S C Lee, S. (2001). Illumination normalization of retinal images using sampling and interpolation. In *Proceedings of SPIE*, volume 4322. (Cited in section 3.2.2.3.)

Wilkinson, C. P., Ferris, F. L., Klein, R. E., Lee, P. P., Agardh, C. D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., Verdaguer, J. T., and Group, G. D. R. P. (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*, 110(9):1677–1682. (Cited in section 2.4.1 and 2.6.)

Williams, G. A., Scott, I. U., Haller, J. A., Maguire, A. M., Marcus, D., and McDonald, H. R. (2004). Single-field fundus photography for diabetic retinopathy screening: a report by the american academy of ophthalmology. *Ophthalmology*, 111(5):1055–1062. (Cited in section 2.4.1, 2.4.2, and 6.1.)

Williams, P. et al. (1989). *Gray's anatomy*. Churchill livingstone Edinburgh. (Cited in section 2.2.)

Wong, T., Klein, R., Couper, D., Cooper, L., Shahar, E., Hubbard, L., Wofford, M., and Sharrett, A. (2001). Retinal microvascular abnormalities and incident stroke: the atherosclerosis risk in communities study. *The Lancet*, 358(9288):1134–1140. (Cited in section 2.4.1.)

Wyszecki, G. and Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae.* New York, NY: John Wiley & Sons., 2nd edition. (Cited in section 2.1.1.)

Xu, J. and Chutatape, O. (2006). Auto-adjusted 3-d optic disk viewing from low-resolution stereo fundus image. *Comput Biol Med*, 36(9):921–940. (Cited in section 7.1.)

Yang, Y. and Pedersen, J. (1997). A comparative study on feature selection in text categorization. In *Machine Learning international workshop 10th conference*, pages 412–420. Citeseer. (Cited in section 6.4.4.)

Zana, F. and Klein, J. C. (2001). Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on Image Processing*, 10(7):1010–1019. (Cited in section 2.2.3, 3.2, 3.2.1.3, and 6.3.)

Zhang, B., Wub, X., Youc, J., Lic, Q., and Karray, F. (2009). Hierarchical detection of red lesions in retinal images by multiscale correlation filtering. *Conf Proc of SPIE*, 7260. (Cited in section 2.3.1, 2.3.1, and 5.3.)