ANDRÉ TEIXEIRA BENTO DAMAS MORA

ADVANCED IMAGE PROCESSING TECHNIQUES FOR DETECTION AND QUANTIFICATION OF DRUSEN

Dissertation presented to obtain the degree of Doctor of Philosophy in Electrical Engineering, speciality on Perceptional Systems, by the Universidade Nova de Lisboa, Faculty of Sciences and Technology.

LISBOA

2010

Acknowledgements

Being the development and writing of this PhD thesis an individual work, due to its complexity and extension, there were people that contributed with new ideas, discussions and motivation, which were very helpful to get this work done. Therefore, I would like to start by thanking all my friends, colleagues, students and researchers that in some way helped me during this work.

First of all, I would like to thank my PhD advisors José Manuel Ribeiro da Fonseca and Pedro Miguel Vieira, for accepting me as their PhD student, for their support, for the enriching discussions, for the motivation, especially when new challenging solutions had to be developed, for always believing in the quality of the work that was done, and for their unconditional friendship.

To José Fonseca, I also have to thank for his sacrifices workwise during my PhD licence, and again it was very enriching to work with him due to his geniality and experience.

To Pedro Vieira I would like to thank him for introducing me to this research topic and by confiding in me the task to successfully complete this work, and the liaisons with multi-disciplinary teams that have enabled me to have a medical and a technical view of the issue.

To Professor Adolfo Steiger Garção in representation of the Electrical Engineering Department and the Uninova – Centre for Technology and Systems, for giving me the conditions to carry on with this work and for his personal support during the this PhD.

A special thanks to *Fundação para a Ciência e Tecnologia* for awarding me a PhD scholarship (SFRH/ BD/ 24719/ 2005), and also for financing the scientific project DRUSAS (POCI/SAU-ESP/57592/2004), which supported the participation on scientific encounters, a scholarship for a young researcher and various materials, which were essential for success of the project.

To my friend Fernando Moitinho, who was also a young researcher financed by *Fundação para a Ciência e Tecnologia*, I would like to leave a special thanks for the contributions to this work, in particular on the design and implementation of *modelling* and *sectioning* algorithms, for choosing me as the advisor in his diploma project, by the professionalism that he puts on every task since the beginning of our professional journey together, and for his friendship.

There are also people from five institutions that were very important to the success of this work, to whom I would like to express my gratitude, and to the institutions they represent. To Ayyakkannu Manivannan from *University of Aberdeen*, for all his collaboration during this work, who introduced to us the problem of drusen automatic detection, supplied retinal images, and who revised some of the articles. To Prof. Monteiro Grillo, MD Carlos Neves, MD David Lopes, MD Ana Ferreira, MD Ana Fonseca from *Hospital de Santa Maria*, for supplying retinal images, for the commitment in analysing the supplied images, and for the partnerships in other projects related to ophthalmologic image analysis. To *Clínica ALM – Oftalmolaser*, in particular to MD Sérgio Ribeiro for analysing several images even during his free-time. To Prof. João Goyri O'Neill from *Faculdade de Ciências Médicas*, for the partnership in the *DRUSAS* project and for their participation on the evaluation studies by analysing retinal images. To MD Simon Brunner from *Rudolf Foundation Hospital, Austria* a special thanks for supplying several image datasets used in this work, for contributing with feedback and requisites to the development of the software applications, for using the developed software in his research activities, and for the joint publication.

A special thank you to Ana Maria by her encouragements and support, and by revising this dissertation, which also helped me to improve my writing in English.

To my family and all the friends who are still there after this last year of social hibernation, I promise to be more present and again share great moments with them. Thank you for the support and encouraging from the beginning until the end of this work.

To my father Zé Manel and in memory of my mother Tatão, I profoundly thank them for the education, the advices, the guidance they gave me, which was fundamental to reach this academic stage, and for my father's comprehension towards my absence.

To Anita, my partner over the last two years, for loving me, for being always there at the good and bad moments, for the comprehension, for the continuous encouragement and motivation specially when the light at the end of the tunnel was dim, for revising this dissertation and many other publications, I am profoundly thankful. I promise now to be less worried and more present!

Abstract

Drusen are common features in the ageing macula, caused by accumulation of extra-cellular materials beneath the retinal surface, visible in retinal fundus images as yellow spots. In the ophthalmologists' opinion, the evaluation of the total drusen area, in a sequence of images taken during a treatment, will help to understand the disease progression and effectiveness. However, this evaluation is fastidious and difficult to reproduce when performed manually.

A literature review on automated drusen detection showed that the works already published were limited to techniques of either adaptive or global thresholds which showed a tendency to produce a significant number of false positives. The purpose for this work was to propose an alternative method to automatically quantify drusen using advanced digital image processing techniques.

This methodology is based on a detection and modelling algorithm to automatically quantify drusen. It includes an image pre-processing step to correct the uneven illumination by using smoothing splines fitting and to normalize the contrast. To quantify drusen a detection and modelling algorithm is adopted. The detection uses a new gradient based segmentation algorithm that isolates drusen and provides basic drusen characterization to the modelling stage. These are then fitted by Gaussian functions, to produce a model of the image, which is used to compute the affected areas.

To validate the methodology, two software applications, one for semi-automated (MD3RI) and other for automated detection of drusen (AD3RI), were implemented. The first was developed for Ophthalmologists to manually analyse and mark drusen deposits, while the other implemented algorithms for automatic drusen quantification.

Four studies to assess the methodology accuracy involving twelve specialists have taken place. These compared the automated method to the specialists and evaluated its repeatability. The studies were analysed regarding several indicators, which were based on the total affected area and on a pixel-to-pixel analysis. Due to the high variability among the graders involved in the first study, a new evaluation method, the Weighed Matching Analysis, was developed to improve the pixel-to-pixel analysis by using the statistical significance of the observations to differentiate positive and negative pixels.

From the results of these studies it was concluded that the methodology proposed is capable to automatically measure drusen in an accurate and reproducible process. Also, the thesis proposes new image processing algorithms, for image pre-processing, image segmentation, image modelling and images comparison, which are also applicable to other image processing fields.

Abbreviations and Notations

AD3RI - Automated Drusen Deposits Detection on Retinal Images
ARMD - Age-Related Macular Degeneration
CV - Coefficient of Variation
FFT - Fast Fourier Transform
FIFO - First In First Out
GAs - Genetic Algorithms
GGD - Generalized Gaussian Distribution
GPL - Gradient Path Labelling
ICC - Intra-Class Correlation Coefficient
IHS - Intensity Hue Saturation
MD3RI - Manual Drusen Deposits Detection on Retinal Images
ODD - Optic Disc Diameters
RGB - Red Green Blue
RIM - Regional Intensity Maximum
RMS - Root Mean Square
ROC - Receiver Operating Characteristic
ROI - Region Of Interest
RPE - Retinal Pigment Epithelium
SAE - Sum Absolute Error
SLO - Scanning Laser Ophthalmoscope
SP - Specificity
SS - Sensitivity
SSE - Sum Square Error

Table of Contents

List of Tables

List of Figures

Chapter 1 -
Introduction

The development of medicine in order to provide longer and better quality of life is probably one of the oldest interests of Man. For a long time has He induced and motivated the development of new treatment and diagnostic techniques aside with the development of new technologies introduced to new equipments, new materials and new computer-aided tools. This has also been the case for Ophthalmology, which has taken care of one of the most important of human senses, the vision, in which new diagnostic equipments, new drugs and others techniques have been constantly developed.

This medical speciality is mainly focused on keeping the human vision functioning correctly; however, the research has shown that the eye and, in particular the retina, is like a mirror showing early signs of many other hidden diseases, like diabetes, hypertension and others, making it even more important for screening purposes. The analysis of the retina is then able to identify several diseases which in some cases don't provoke vision disorders but are early signs of other pathologies.

It has been world widely observed a growing incidence of vision disorders, especially on Age-Related-Macular-Degeneration (ARMD), which has been reported as the main cause of vision loss in the developing countries (World Health Organization 2007). It is mainly caused by life-style changes and its symptoms should be visible in a retina examination. Up to now, it has been diagnosed visually by the Ophthalmologist who follows specific guidelines for its assessment. This task is fastidious, subjective and lacks reproducibility. Nevertheless, the use of computer-aided tools for this diagnosis is still incipient. The use of image processing techniques for automating the analysis of these images will certainly increase the

process reproducibility and provide quantitative data to the Ophthalmologist, enabling him to focus his attention on the diagnosis.

This work has dedicated especial focus on ARMD and in particular on drusen which are one of its early signs. These are yellow spots which appear in the retina surface and can vary in size depending on their type and disease evolution. For the ophthalmologists the quantitative evaluation of the total area affected by this disease in a sequence of images taken during a long term treatment will positively help on assessing its effectiveness. Currently, to assess it, they need to manually mark drusen in the retinal images and quantify these areas. This is a procedure that requires attention, precision and time from the ophthalmologists, which is incompatible with their daily interventions. The subjectivity of the criterion to detect drusen also affects the reproducibility of the technique and increases the variability among ophthalmologists analysis.

This thesis proposes a methodology which is able to automatically process retinal images and quantify the area affected by drusen, using digital image processing techniques. The complexity of these images along with the low contrast and non-uniform illumination make the development of these techniques a difficult task. The work is a contribution to both medicine and engineering fields, in the sense that the tools developed will certainly improve the clinicians' work by freeing them from the manual markings and the engineering field which will benefit from the developments of new image processing algorithms, as well as a case-study for other biomedical applications of image processing.

The next subsections will introduce the reader to the anatomy of the human visual system, to the pathologies which are related to the retina and to the diagnosis and treatment of ARMD.

1.1 - The human vision

Of the five human senses, the eye is one of its most important sense organs by the

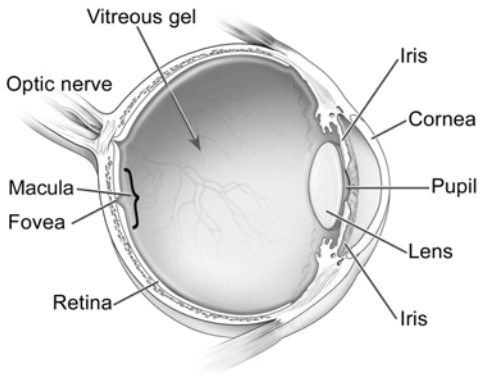


Figure 1.1 – Cross-sectional view of a human eye. (from (National Eye Institute 2009)).

quality and precision of the signals it captures and by the reduction of quality of life within its absence. It was estimated that 70% of the sensorial information interpreted by humans is captured by the eye (Davidovits 2001). It has the ability to capture the light and convert it in the retina into electric signals that are sent to human brain for interpretation. Its functioning is similar to conventional image capture systems due to its lenses that refract and focus the incoming light in the sensorial region.

The optical system (figure 1.1) is composed by the lens, the iris and the cornea. As a result of the incoming light and in order to refract and focus, the lens reshapes with the help of auxiliary muscles. The control of the amount of light that enters the eye through the pupil is done by the iris, which is a tissue that is able to contract and expand, by decreasing or increasing the size of the pupil, respectively (figure 1.2.a). The cornea is a transparent and protective layer that covers the iris and the pupil and is the first refracting layer. With approximately 43 dioptres of refractive power the human cornea is the major focusing element of the eye, although it is fixed. The variable focus is obtained by the lens which in a natural environment has approximately 18 dioptres of refractive power (figure 1.2.b).

The light that enters the eye is projected into the retina, which contains the photoreceptors. There are different types of photoreceptors, some are for low-light vision and black-white perception (rods) and others are for colour perception and daytime vision (cones). Although these are dispersed along the retina, which occupies 72% of a sphere with 22mm of

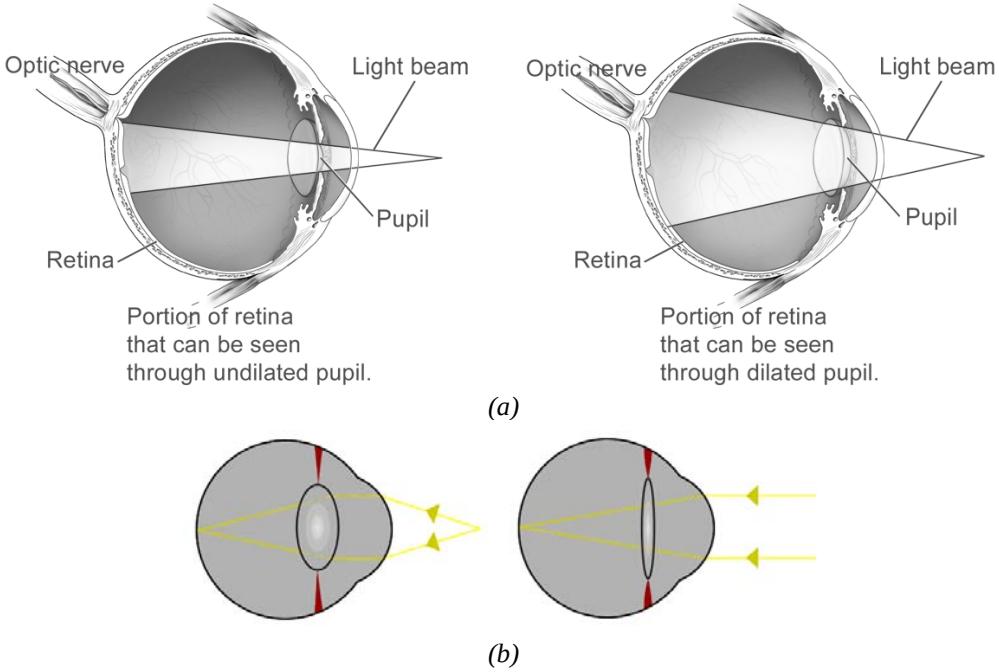


Figure 1.2 – Light propagation in the retina. (a) The light behaviour for an undilated and a dilated pupil; (b) Light refraction in the eye with a relaxed and a stretched lens.

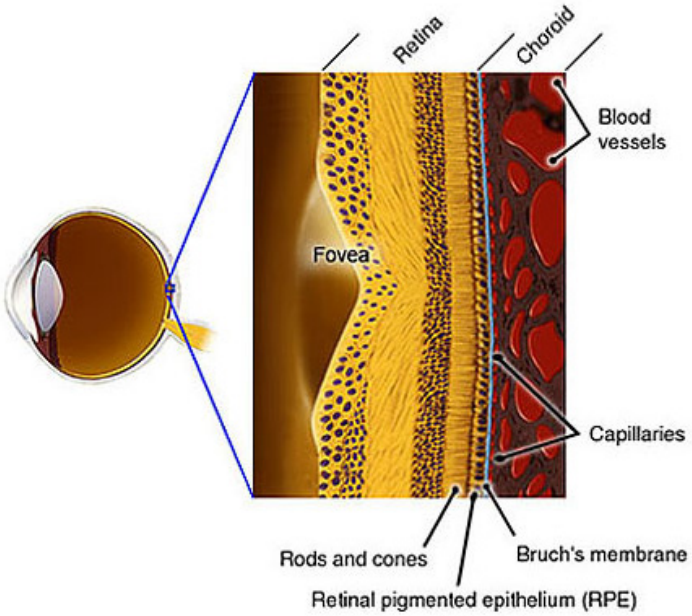


Figure 1.3 – Retina Layers distribution. (*from* (American Health Assistance Foundation 2009))

diameter, they are more concentrated on the central part of the retina, the macula. This latter is about 6000µm of diameter and contains a high density of photoreceptors. The photoreceptors that are outside the macula capture the peripheral vision, which in case of damage is less noticed or even unnoticed than in the macula. In the centre of the macula with 1000µm of diameter is the fovea. It contains 50% of the photoreceptors of the retina being responsible for the sharp and high resolution central vision used by humans for reading, watching television, driving, and any other activity where visual detail is of primary importance.

The retina is divided into layers from the nerve fibre layer to the Bruch's Membrane. The light that reaches the retina crosses the nerve fibre and the ganglions' layers to be captured by the photoreceptors (rods and cones). These return the light information to the ganglions which gather and compress the information from several neighbouring photoreceptors and send it through the nerve to the brain. The Retinal Pigmented Epithelium (RPE) is a layer of cells that protects and nourishes the retina, removes waste products, prevents new blood vessel growth into the retinal layer and absorbs light not absorbed by the photoreceptor cells; these actions prevent the scattering of the light and enhance vision clarity. Finally, the Bruch's Membrane is a thin layer that acts as a blood-retinal barrier and as a support to both RPE and Choroid.

The information collected by the retina is gathered into a set of nerve fibres which are forwarded to the brain through the optic nerve. This latter includes also the veins and arteries which supply the blood to the eye. The optic disc is located within the retina and does not have photoreceptors which create a blind spot in the visual field.

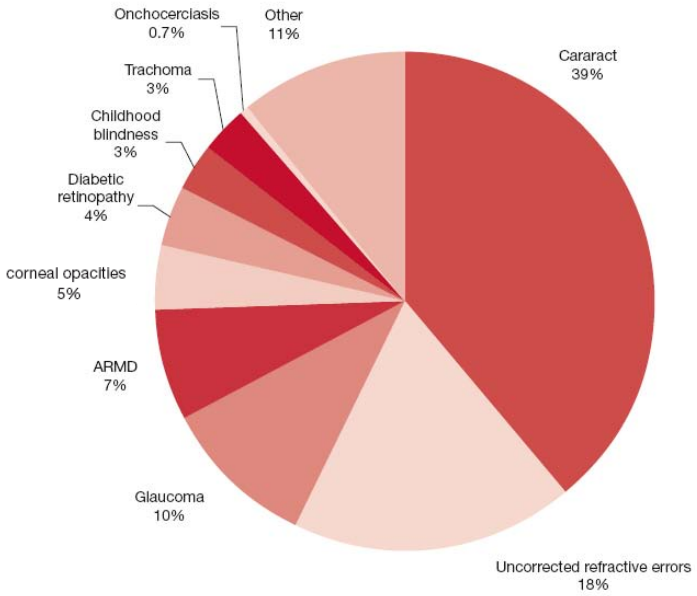


Figure 1.4 - Global causes of blindness due to eye diseases and uncorrected refractive errors estimated in 2006. (*from* (World Health Organization 2007))

1.2 - Introduction to Retinal Diseases

Blindness is one of the most undesirable consequences of any eye disease. According to the World Health Organization (2007) the major causes of blindness that occur nowadays worldwide are the opacity of the lens provoked by Cataracts, Glaucoma which affects the optic nerve and involves the loss of ganglions, and Retinopathies that affect the retina, such as ARMD and Diabetic Retinopathy by distorting the image, or in an advanced stage, with the loss of photoreceptors (figure 1.4).

In this work only a particular retinopathy is approached, the ARMD. However, in order to contextualize the reader an overview on other retinopathies which can have confounding characteristics such as the Diabetic Retinopathy and the Ocular Melanoma will be presented in the following sections.

1.2.1. Diabetic Retinopathy

A patient with Diabetes may experience several ocular complications, such as Cataracts, Glaucoma and Diabetic Retinopathy, which can lead to irreversible blindness. This latter is the most common within the patients with diabetes and is one of the leading causes of American Adults blindness (Prevent Blindness America 2008). It is caused by changes in the blood vessels of the retina that can lead to haemorrhages, blurring the vision. These changes can be new vessels that grow on the surface of the retina or vessels that swell and start bleeding.

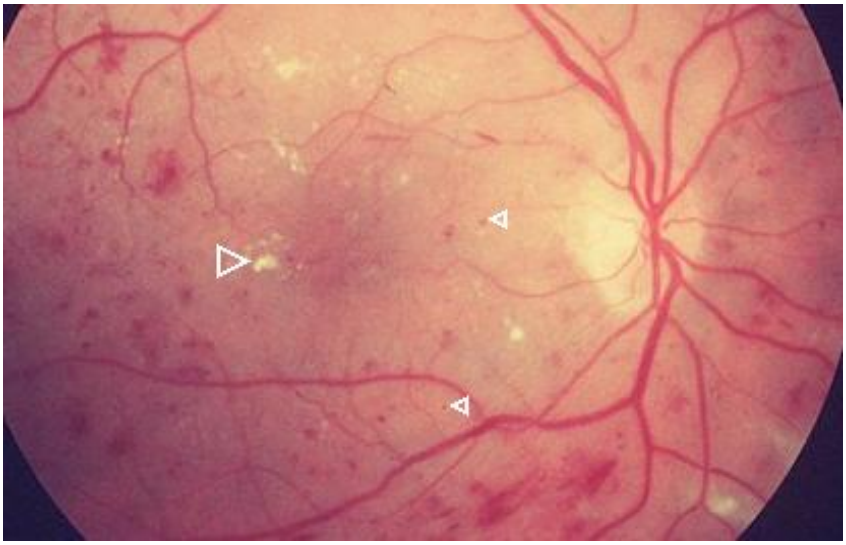


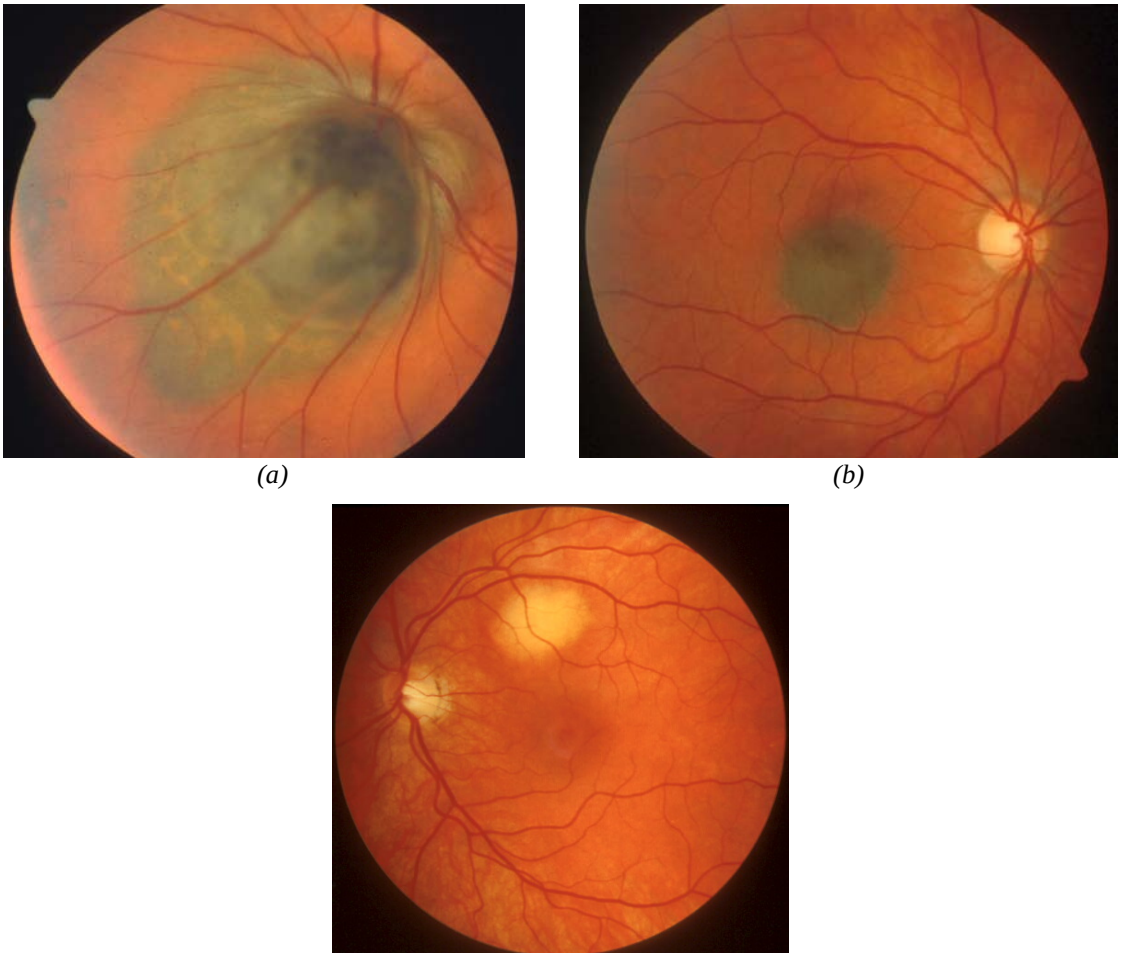
Figure 1.5 – A retina affected by Diabetic Retinopathy, where cotton wool spots (bright spots) and haemorrhages (darker spots) are visible.

This disease often has no early signs. Only after small haemorrhages do the patients feel their vision blurred. Firstly this sensation can be temporary. However, after some days or weeks, a severe bleeding is likely to occur and damage the retina irreversibly. In a funduscopic examination of a retina affected by Diabetic Retinopathy, the ophthalmologist may find cotton wool spots and haemorrhages (figure 1.5). The cotton wool spots are regions where the blood supply has been obstructed, exhibiting a white reflection in a distorted region as a consequence. The haemorrhages are the darker spots which exhibit an irregular shape, as illustrated in figure 1.5.

1.2.2. Ocular Tumour

Ocular Tumour is a rare disease which can appear either in the iris or in the Ciliary Body or in the choroid. According to Liverpool Ocular Oncology Centre (2005), within the various types of tumours the most commons are Uveal Melanoma, Naevi and tumours metastasis. Those affecting the retina are the ones which grow in the Choroid. Most of these tumours remain unnoticed until they are detected in a routine retinal examination. However, in an advanced stage they affect the patient's vision, leading in many cases to irreversible blindness and ultimately become fatal.

The Uveal Melanoma is a malignant tumour that develops in the three eye structure specified above. Its consequences are a visual field deficit, pain and eventually the loss of vision. It is also common to develop metastasis which can be fatal to the patient. Most patients with Choroidal Melanoma have no symptoms at all, only detected in a routine fundus examination appearing with a different pigmentation (figure 1.6.a).



(a) *(b)*

(c)

Figure 1.6 – Examples of retinas showing signs of Ocular Tumours. (a) ocular melanoma (advanced stage); (b) Naevus; (c) Metastasis. (*from* (Liverpool Ocular Oncology Centre 2005))

Naevi tumours are benign tumours which present no health risks unless they develop into malignant tumours which are then classified as Melanomas. They are from the same class as the skin mole, being characterized in fundus images by a darker region (figure 1.6.b). This coloration is due to a high concentration of the pigmenting agent, melanin. The patients usually do not have symptoms, being also only detected on a retinal examination.

Other class of ocular tumours are the metastasis of other malignant tumours, which can develop in the eye growing rapidly and leaking large amounts of fluid which cause progressive loss of vision. They usually appear as large sized spots in yellow or white colour, as it is show in figure 1.6.c.

1.2.3. Age-Related Macular Degeneration

The ARMD is the most common cause for irreversible blindness in the developed countries (Friedman, O'Colmain et al. 2004; AMD Alliance International 2006; Cook, Patel et al. 2008). According to the World Health Organization it is the third cause of irreversible

blindness. It is a disease that causes progressive damage to the macula, leading to a distorted vision and ultimately to a complete loss of central vision. Peripheral vision is usually retained in ARMD and therefore total blindness (inability to see all light) does not typically occur even in advanced macular degeneration.

As the name suggests ARMD is age related, common among people aged 55 years and older. Many epidemiologic studies were and are being carried out to evaluate the causes, the risk factors and its prevalence in some races, genders or parts of the globe (Vingerling, Hofman et al. 1996; Pagliarini, Moramarco et al. 1997; Curcio, Medeiros et al. 1998; Frank, Puklin et al. 2000; Oshima, Ishibashi et al. 2001; Jonasson, Arnarsson et al. 2003; Varma, Fraser-Bell et al. 2004; Augood, Vingerling et al. 2006; Wang, Rochtchina et al. 2007; Chen, Cheng et al. 2008; Xu, Li et al. 2008). One of the conclusions taken from these studies was the profile of the patients that have higher probability of developing ARMD, which is:

- Age - signs are present in about 14% of people aged between 55 and 64, 20% aged between 65 and 75 and up to 37% of people over 75 (Wang, Mitchell et al. 2007).
- Gender - more common with women than with men. This partially may be explained by the fact that women live longer than men (Klein, Klein et al. 2002; Wang, Mitchell et al. 2007).
- Race - more common in Caucasians than other races. This partially may be due to the pigment in the eye or eye colour. It may also have to do with differing diets and sun exposure (Schachat, Hyman et al. 1995; Frank, Puklin et al. 2000; Varma, Fraser-Bell et al. 2004).
- Eye Colour - more common with people with blue eyes. Although, there is no strong association, this may be related to damage associated with exposure to ultraviolet light (Frank, Puklin et al. 2000).
- Genetics – It was found that a genetic factor is associated with ARMD. If someone in the family contracted ARMD it is likely that some of their descendents may develop it also (Frank, Puklin et al. 2000; Smith, Assink et al. 2001; Haddad, Chen et al. 2006).

These studies found also risk factors which can be the genesis of ARMD. Most of these are due to people's habits, which can be avoided to reduce the probability of developing ARMD. Some of the identified factors were:

- ARMD in One Eye – If one of the eyes is affected with ARMD the probability to develop in the other eye is higher (Sandberg, Weiner et al. 1998).

- Cardiovascular problems – It was found a higher incidence of ARMD on people having hypertension (Tan, Mitchell et al. 2007; Cukras and Ferris 2008).
- Exposure to Sunlight – The long exposures to sunlight, especially to ultraviolets without protective sunglasses are related to a higher incidence of ARMD. The exposure to blue light has also been identified as an increase factor (Fletcher, Bentham et al. 2008).
- Dietary habits - The dietary habits may also influence negatively or positively the risk of ARMD. It was found that a fat dietary increases the risk, while a dietary based on fish reduces it (Seddon, George et al. 2006).
- Smoking – Smoking is being highly correlated with the presence of ARMD. It has been consistently confirmed in several studies that the risk is two to three times higher in smokers than non-smokers (Vingerling, Hofman et al. 1996; Delcourt, Diaz et al. 1998; Seddon, George et al. 2006).
- Drusen – These are small lipid deposits that appear in retina surface and can lead to ARMD. Alone they do not provoke macular degeneration, but contribute for its development (Pauleikhoff, Barondes et al. 1990).

The patient's symptoms are a distorted vision and/or occlusions in the formed image, as exampled on figure 1.7. If untreated, these symptoms will increase progressively until a total occlusion. It is characterized by a detachment of the macula from the Choroid, which contributes to a deficient nutrition of the macular cells causing their death progressively. This detachment can be originated from one of two forms of ARMD: *wet* or *dry*.

The *wet* form is less common but more severe and is caused by choroidal

Figure 1.7 – Visual perception from a patient with ARMD.
(*from* (St. Luke's Cataract & Laser Institute 2008))

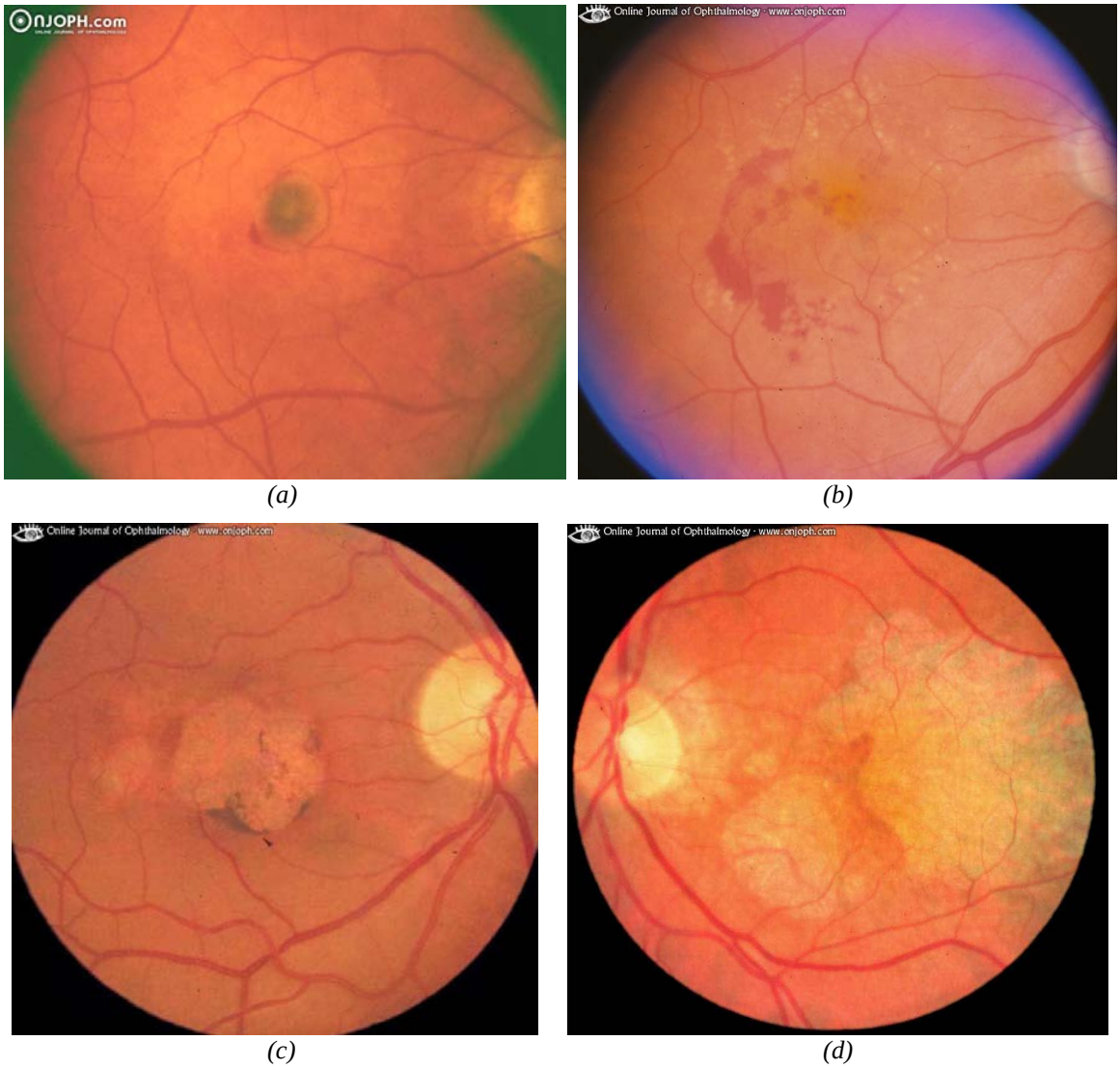


Figure 1.8 –Examples of: retinal abnormalities. (a,b) Choroidal Neovascularisation. On the right-side image there is a mixture between neovascularisation and drusen; (c,d) Geographic Atrophy.
(*from* (Machemer, Michelson et al. 2009))

neovascularisation, i.e., the development of abnormal vessels beneath the RPE layer. These can provoke macular detachment and bleed causing profound loss of vision. It usually appears in fundus images as darker regions due to the bleeding (figure 1.8 (a,b)).

The dry form is characterized by the presence of drusen which can degenerate and grow provoking a gradually distorted vision and occlusions. When they appear outside the macula, in the peripheral vision, the brain compensates the failures with the neighbours' photoreceptors not being noticed by the patient. Its last stage is the Geographic Atrophy which is identified as a yellow and brighter region with a size larger than 175µm with sharp edges (figure 1.8(c,d)). The first stage where the drusen are visible, which is the core of this work, will be deeper explained in the next section.

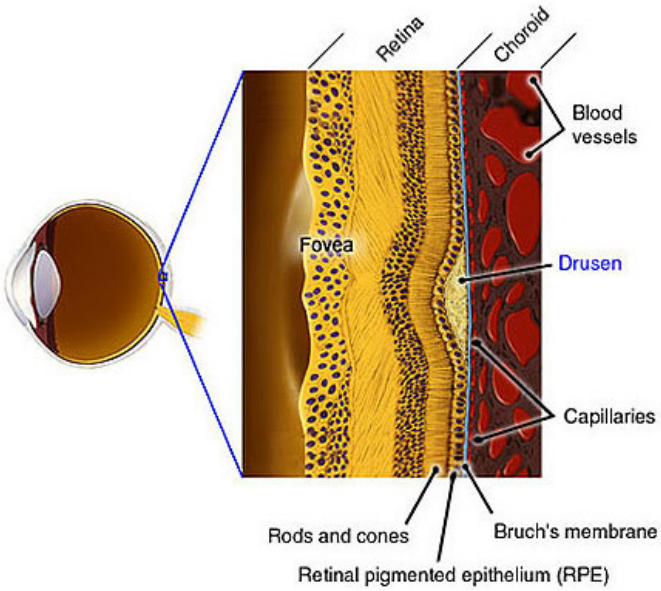


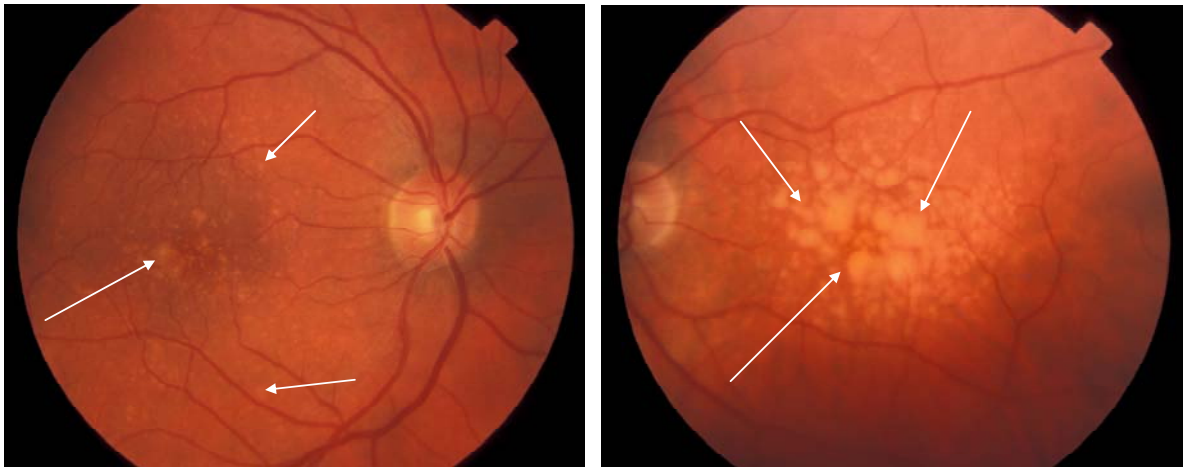
Figure 1.9 – The appearance of drusen between the Bruch's membrane and the RPE.
(from (American Health Assistance Foundation 2009))

1.2.4. Drusen as a risk factor for ARMD

The drusen are deposits of extracellular materials in retina between the RPE and the Bruch's membrane (figure 1.9) and they are considered a strong risk factor for ARMD (Pauleikhoff, Barondes et al. 1990; Hageman, Luthert et al. 2001). They have also been related to the appearance of retinal detachments, which provokes loss of vision (Hartnett, Weiter et al. 1992; Bonanomi, Maia et al. 2006). By separating the two retinal layers the drusen will make it difficult to transfer the nutrition and oxygen from the Choroid to the RPE which potentially leads to the degeneration of some photoreceptor regions (Johnson, Lewis et al. 2003).

Drusen deposits can appear scattered or concentrated, depending on the stage of ARMD. In the early stages the accumulation and size of drusen is usually small. In these cases as the human brain compensates the small size abnormalities, the patient feels little or no impact on his vision. However, when it degenerates in ARMD, the drusen deposits grow in accumulation and size developing symptoms such as blurred vision. The patients will generally require more light and will often indicate a loss of vision that is more pronounced in the centre of the eye.

Drusen appear in fundus images as yellowish spots in or around the macula. Clinically, they can be divided in two types: the hard drusen and the soft drusen (Bird, Bressler et al. 1995). Hard drusen (figure 1.10.a) are mainly characterized by small sized



(a) *(b)*

Figure 1.10 – Retinal images containing: (a) hard drusen; (b) soft drusen.

spots with less then 63µm of diameter and sharp edges, while soft drusen (figure 1.10.b) are medium sized spots between 63µm and 250µm of diameter with smooth edges, which can be confluent with irregular shapes. They also provoke elevations on the retina surface, as consequence of the accumulation of deposits beneath the retina.

The detection of drusen in early stages and the control of its progression is therefore important to reduce the risk of degeneration and consequently the development of ARMD. There are several possible treatments and diagnosis techniques which will be presented next.

1.3 - Diagnosis, Follow-up and Treatment

Many of the pathologies that affect the retina can be visible in retina surface, hidden in a deeper layer or appear in other invisible forms. Thus, some of these pathologies require different complementary diagnostic exams to correlate information between them and produce a reliable and well supported diagnosis. In this work it is analysed the most common diagnostic techniques for drusen detection, which are the Amsler Grid and retinal imaging. The biogenesis, causes or effects are not analysed, although its main goal is to study the progression of the drusen and evaluate the treatment efficiency.

Further on it will be presented an overview of the Amsler Grid and the imaging techniques used in retinal diagnosis which can be used to detect and evaluate drusen. As contextual information for the reader the available drusen treatments will also be presented.

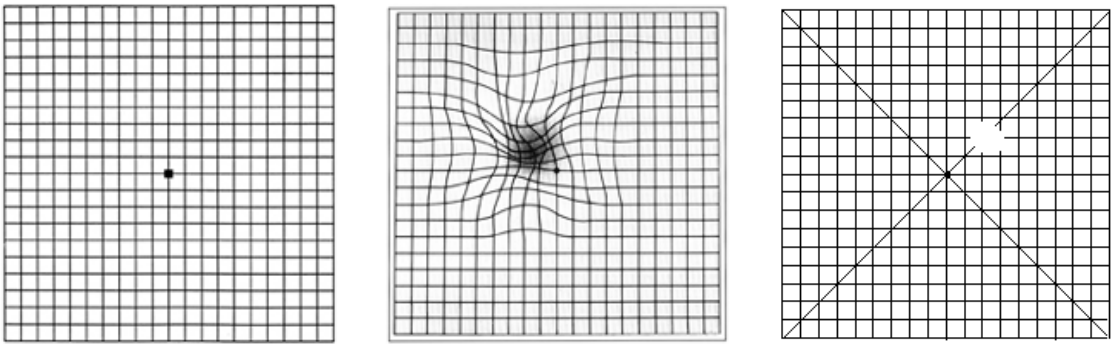
1.3.1. Amsler Grid

A simple method for early detection of ophthalmologic abnormalities is a self-test to detect changes in the visual perception of the surrounding environment. With a monocular

inspection of straight lines, such as, door frames, lines of text, buildings edges and others, visual acuity changes can be detected if these lines present signs of distortion. This is called environmental Amsler (Fine, Elman et al. 1986; Amsler Grid 2010), in relation to the Amsler Grid, which is also a visual acuity exam.

The Amsler Grid (figure 1.11.a), a grid of horizontal and vertical lines with a small dot at the centre, was developed in 1945 by Marc Amsler (a Swiss Ophthalmologist) and is a non-invasive and low-cost diagnostic tool to analyze visual disturbances caused by changes in the retina. A patient with macular disease may see wavy lines and some lines may be missing, as showed in figure 1.11 (b,c).

The test begins when the ophthalmologist asks the patient to look at the grid, centred on the small dot, at a predefined distance (14 in.), with an eye at a time, followed by the description of the image they are seeing. If they see distorted lines or missing ones, these are probably signs of ARMD. However, it is not very accurate, being required further diagnostic techniques.



(a) (b) (c)

Figure 1.11 – Amsler Grid. (a) normal grid; (b,c) visual perception by a person with ARMD.

1.3.2. Imaging equipments

For the retina examination there are retinography equipments to capture images from its surface, for tomography and topography analysis. The first retinal in-vivo examinations were made in 1851 with the invention of the ophthalmoscope by Hermann Von Helmholtz. Since then with optics and fast electronics developments the current equipments can capture and store on computer high-quality retinal images without the need of pupil dilation as it was before. These equipments can capture images from the retina surface layers and deeper layers, giving ophthalmologists important information which is used to elaborate patient diagnosis.

In the following subsections it will be presented the available techniques to study the retina. The mean of some of them is beyond the scope of this work which is to apply image

processing algorithms to images captured from the retina's surface. However, they will be presented as a contextual perspective.

1.3.2.1 Direct Ophthalmoscope

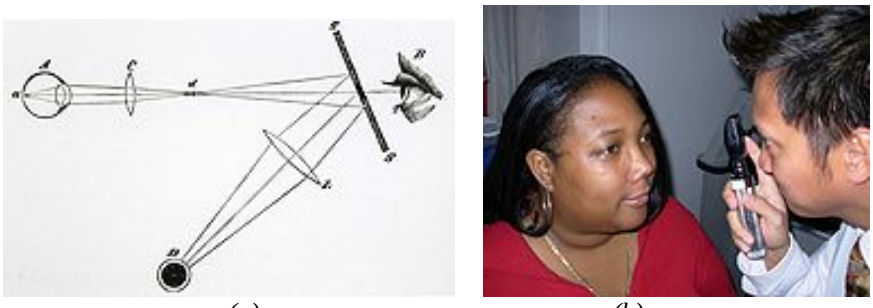
The first equipment used for retinal examination, which is still widely used, is the ophthalmoscope. It was originally invented by Charles Babbage in 1847, although Hermann Von Helmholtz in 1851 reinvented it to what it is nowadays (Helmholtz 1851). Helmholtz showed that when light is transmitted into the pupil it is reflected in the retina and follows the same path when entering the pupil but in opposite direction, and he developed a prototype in which he could see the retina surface.

The ophthalmoscopes are handheld equipments that illuminate the retina, used by observers that place themselves in the light path, between the retina and the lamp to view the light reflected on the retina's surface (see figure 1.12 (a,b)). To accomplish this process, a set of condensing lens conduct the light into the pupil with a semi-transparent mirror. The mirror reflects the light from the lamp into the pupil being semi-transparent in order to enable the observer to see the reflected light from the retina. Besides their conducting objective, they are also used to correct the refracting errors of the patient or the observer and to magnify the image.

This equipment is used for routine examinations, because it is practical and portable, being a very important instrument for early detection of retinal abnormalities. However, the devices that can capture the image for further processing in the computer are not so practical and are less common.

1.3.2.2 Indirect Ophthalmoscope - Fundus cameras

The fundus is the interior surface of the eye on the other side of the lens. It contains the retina, the optic nerve and the macula, and can be observed through the lens. A fundus



(a) *(b)*

Figure 1.12 – Direct Ophthalmoscopes. (a) the ophthalmoscope setup. (from (Helmholtz 1851)) (b) ophthalmoscopy examination.

camera or retinal camera is similar to an ophthalmoscope which is able to view the fundus and record this observation either on analogue or digital media.

The optical principle is similar to the ophthalmoscope (figure 1.13) although the observation light (6) is focused on the pupil through a series of lenses and apertures to form an annulus light (7). The reflected light (5) follows a different path passing through the centre of the annulus. Since the light paths of the two lights are independent, there are minimal reflections of the light source captured in the formed image. It has two different light sources to illuminate the retina, one with less intensity used for the alignment (6) and another with a flash light (8) to capture the image. When the observer presses the capture button the mirror in front of the camera (2) is lifted, the flash light (8) is fired and the image is captured (3).

The camera can be analogue or digital. Most retinograph setups are connected to a monitor and professional printer; only the recent ones have a digital output and are directly connected to a computer. The normal procedure is to digitalize the 35mm slide or, in some cameras with analogue output, to capture these same images using an image acquisition board.

There are two distinct versions for these retinal cameras. Those that require pupil dilation, which are the *mydriatic* retinal cameras (see figure 1.14.a), and those that don't, the *non-mydriatic* cameras (see figure 1.14.b). When the eye is being illuminated for alignment the pupil contracts decreasing the field of view and limiting the amount of light that can illuminate the retina. This is the reason why the pupil must be dilated during the image capture for the *mydriatic* retinograph. The *non-mydriatic* retinograph uses infrared light to illuminate the retina during the image alignment, similar to darkness, which stimulates natural

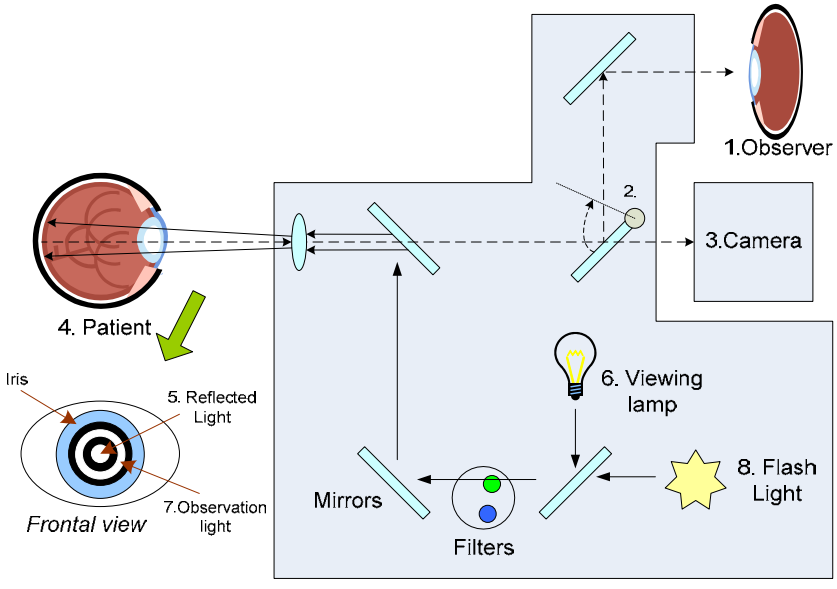


Figure 1.13 – Fundus camera diagram.

(a) *(b)*

Figure 1.14 –Fundus camera examples. (a) Canon CF1 Digital mydriatic retinal camera; (b) Canon CR1 Mark-II Digital non-mydriatic retinal camera. (from (Canon 2009)).

pupil dilation. Notice in the figure 1.14 that the *mydriatic* uses optical viewer as the *non-mydriatic* uses a LCD to display the retina illuminated by infrared light. However, some studies (Lim, LaBree et al. 2000) revealed that the *non-mydriatic* retinographs have a low sensitivity rate and a high specificity rate, being clinically less useful than standard dilated 35-mm fundus slide images.

These retinographs are equipped with light filters to obtain images with different characteristics. One of such is the red-free filter that is used to obtain higher contrasted images, by enhancing the visibility of the retinal vasculature, haemorrhages, drusen and exudates. To obtain angiographies the filter is changed to a band-pass of blue light or infrared depending on the fluorescent dye used.

In angiography the patient receives an intravenous injection of a fluorescent dye which enhances the vessels contrast over the background. The method starts by illuminating the retina with a specific wavelength for the dye to fluoresce on another wavelength. Then, by filtering the illuminating wavelength, the fluorescent one passes producing thus very high-contrast image of vessels. The capture of a sequence of photographs of the progression of the dye into the vessels which reveals the flow dynamics and related pathologies is one of the main diagnostic exams. The two most common angiography methods use *sodium fluorescein* dye, which emits a yellow-green light (*wavelength* = 530nm) when illuminated by a blue light (*wavelength* = 490nm) or an *Indocyanine green* dye (ICG) which absorbs and emits light in the near infrared spectrum (*wavelength* 805nm and 835nm, respectively). The ICG is a method which captures an image of the choroid, which can be useful for some clinical purposes. In figure 1.15 it is shown some of the differences between these two dyes.

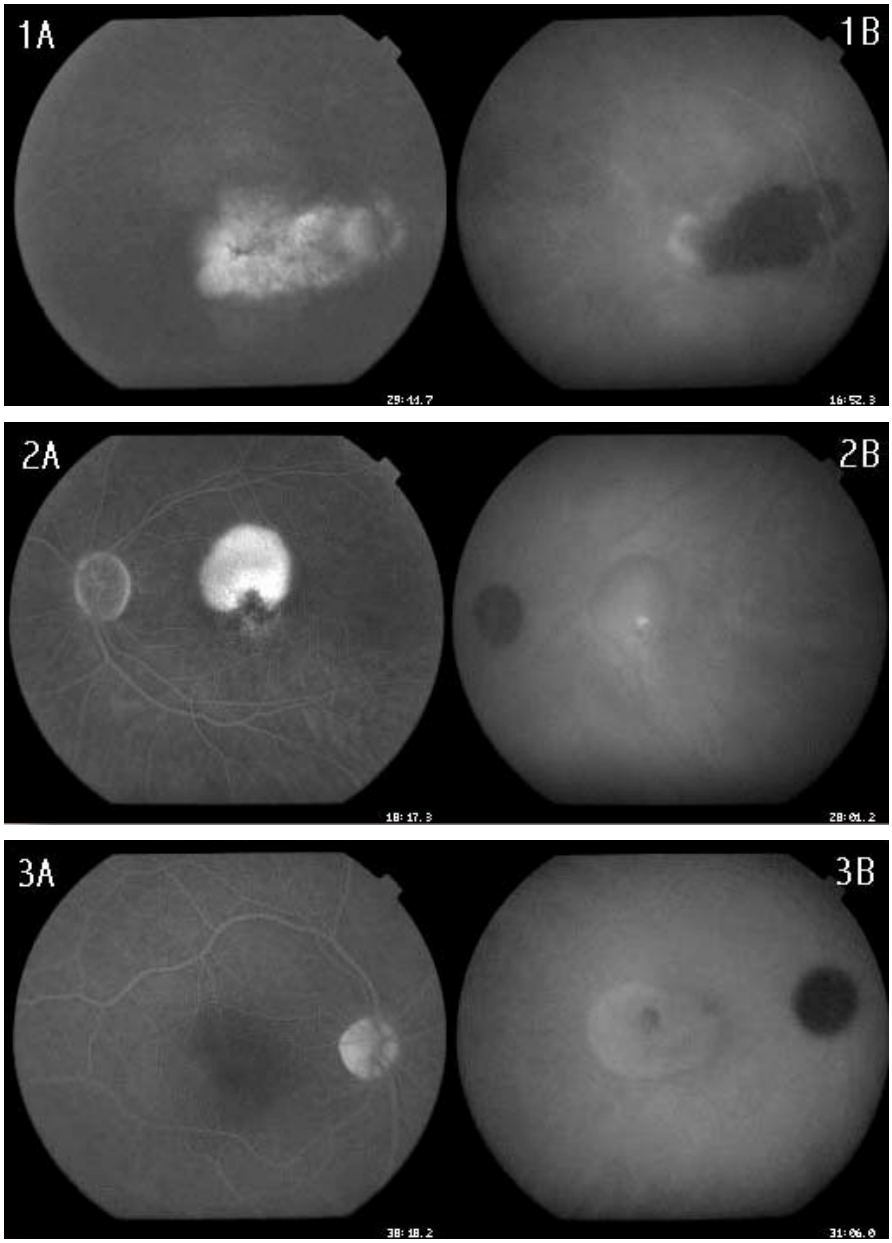


Figure 1.15 – *Fluorescein* (A) and *Indocyanine* (B) angiograms side-by-side comparison. (from (Miller 2006))

According to Owens (1996) the physical properties of ICG make it potentially useful for the assessment of ARMD. However, the theoretical advantages are not verified in the practice. There were few studies referring drusen assessment in angiographs, one of such evaluated 37 patients (Scheider and Neuhauser 1992) where drusen were hypofluorescent 70% and hyperfluorescent only 22%. This showed that angiography images were not suitable for drusen detection.

1.3.2.3 Scanning Laser Ophthalmoscope

A Scanning Laser Ophthalmoscope (SLO) is a retinograph which uses a narrow beam of light from a laser to scan the retina in a raster fashion. The light reflected from the spot is

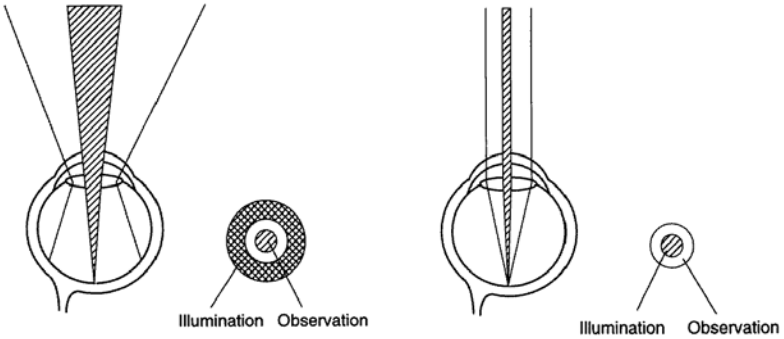


Figure 1.16 – Retinal illumination diagrams. (a) fundus camera (b) SLO.
(*from* (Sharp and Manivannan 1997))

detected and synchronously decoded to form an image on a monitor. This process takes few milliseconds to complete a single image scan that avoids image blurring due to eye movements. This is a low power laser to avoid damaging the retina (Sharp and Manivannan 1997).

To illuminate the retina, the SLO uses a different approach than the fundus cameras. Since the reflected light is less powered than the illumination, this technique maximizes the capturing area using a small aperture in the central part of the pupil for the laser beam and uses the remaining area to collect the reflected light (figure 1.16). With this technique the efficiency is improved, allowing a low powered light to be used.

The functioning is similar to the previous ophthalmoscopes where the key-point is to transport the light one way and the reflected light the other way, through the same path

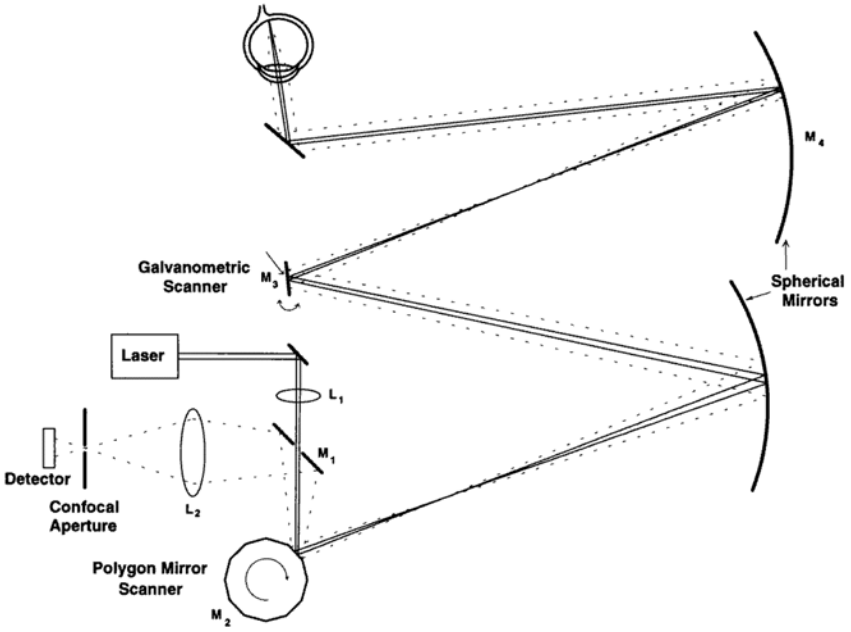


Figure 1.17 – An optical diagram of the SLO.(*from* (Sharp and Manivannan 1997))

without interferences. Here the illumination is easier to mask, since its diameter is narrower. The diagram (figure 1.17) presents a generic schema for SLO equipments.

The equipment's purpose is to focus the laser on the retina, by passing it through the lens (*L1*) and the mirror aperture (*M1*), which is responsible for separating the illumination from the reflected light. The laser is then deflected horizontally and vertically using an optical system (*M2*, *M3* and *M4*) to scan the retina surface. The reflected light follows the same path back to the equipment and is forwarded by the mirror (*M1*) to the focusing lens (*L2*) and then to the *photo-detector* before crossing the *confocal aperture*.

The SLO can be used on one of three possible modes: *direct* or non-confocal, *confocal* and *indirect*, depending on the size and position of the confocal aperture. In the *direct* mode the confocal aperture is larger when compared to the laser beam width (figure 1.18.a). In this mode the scattered light that is reflected from the retina is also captured by the photo-detector. The resulting images have a high overall brightness as the light reflected from all retina layers is captured.

In the *confocal* mode (figure 1.18.b) the aperture size is reduced in order to filter only the direct reflected light. By adjusting the distance between the *confocal aperture* and the *photo-detector* it is possible to focus the receptor to one of the retina layers (figure 1.18.d and figure 1.19.a). This enables topographic imaging to be performed.

Finally, the *indirect* mode (figure 1.18.c) uses a large *confocal aperture* with a central

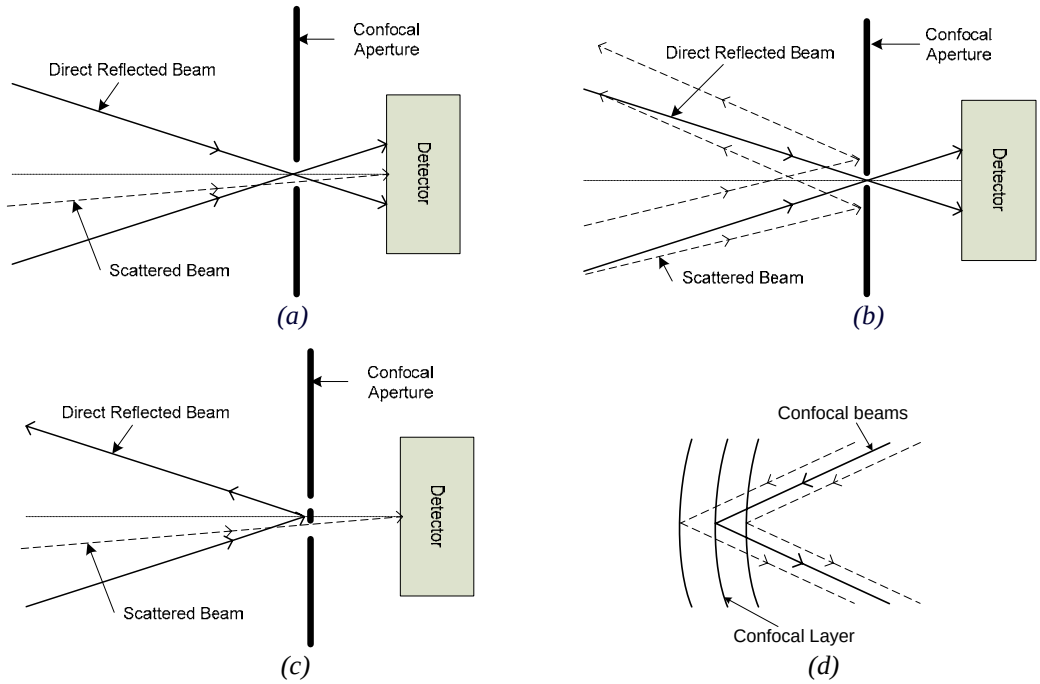
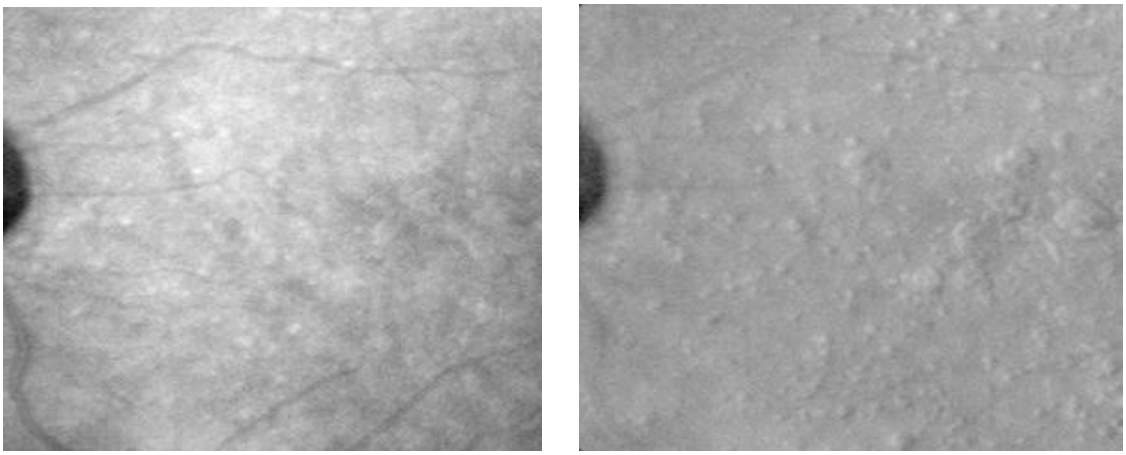


Figure 1.18 – Diagram for an SLO capturing in: (a) direct mode; (b) confocal mode; (c) indirect mode; (d) confocal retina layers. (*from* (Sharp and Manivannan 1997))



(a) *(b)*

Figure 1.19 – Retina images containing drusen taken with SLO. (a) confocal mode; (b) indirect mode. (*from* (Aberdeen-Biomedical-Imaging-Centre 2009))

stop to cut the direct reflected beams and capture only the scattered ones. This method has been found to be useful in taking drusen images (Manivannan, Sharp et al. 1993) as illustrated in figure 1.19.b.

One of the limitations of these techniques is that the image is illuminated with a single colour, producing monochromatic images. To overcome this issue, some authors (Manivannan, Kirkpatrick et al. 1998; Vieira, Manivannan et al. 2002) have improved the conventional SLO to produce true colour images of the fundus. Their method endorsed a retina illumination with a three laser colour (blue, green, red) sequence and combined them to

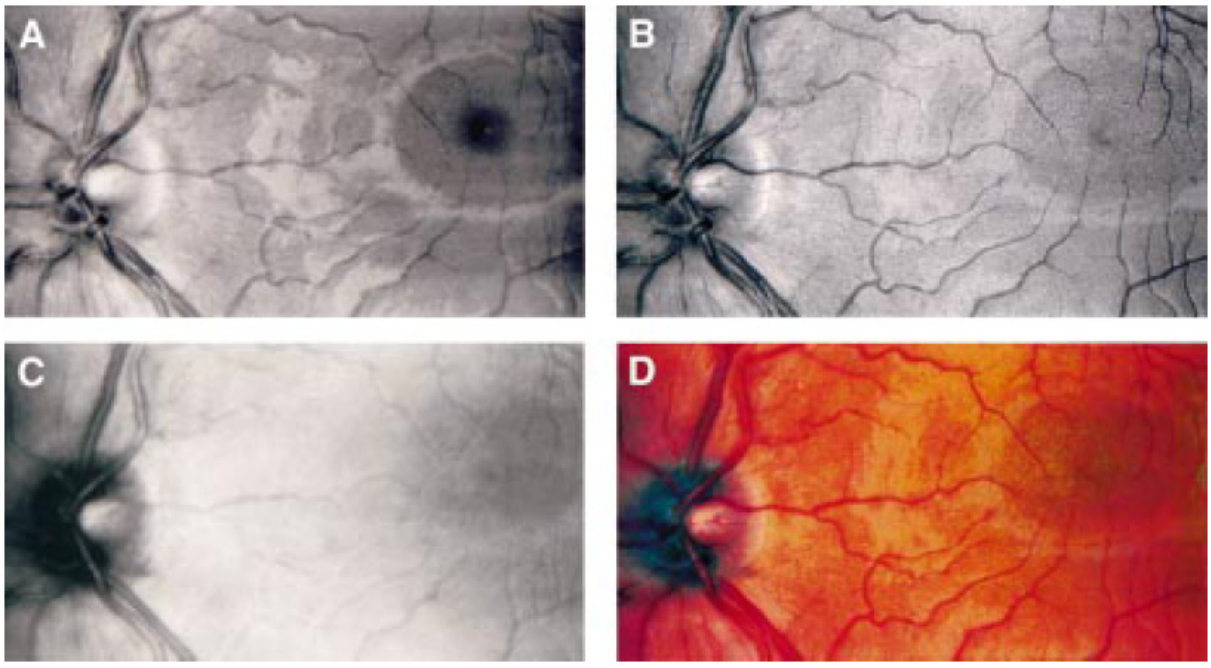


Figure 1.20 – Monochromatic retinal images obtained with laser beams: (a) blue; (b) green; (c) red. (d) The generated true colour image. (*from* (Manivannan, Kirkpatrick et al. 1998))

produce one true colour image (RGB). The image characteristics are slightly different from fundus retinographs, but with the advantage to photograph deeper layers of the retina. In figure 1.20 (a,b,c) the three monochromatic images taken with blue (488nm), green (523 nm) and red (670 nm) lasers are presented, as is the combined colour image in figure 1.20.d.

1.3.2.4 Optical Coherence Tomography

The Optical Coherence Tomography (OCT) is a non-invasive imaging equipment which performs three-dimensional depth scans of the retina. The first *in vivo* two-dimensional depth scan using OCT in ophthalmology was presented by Fercher (1990). In the following years it was further developed achieving micrometer resolution and cross-sectional imaging capabilities. It is able to capture images from a depth of 1 to 2 mm below the surface in the retina.

The OCT is based on interferometry[1] techniques to analyze the reflection from the retina surface. Typically an infrared light source (800nm or 1.3µm), which provides a deeper penetration in the tissues with a low scattering, is used. In (Wang, Nelson et al. 2003) the relation between the resolution of OCT and the light source wavelength was studied. It was shown that to reduce the dispersion caused by the water in the tissue, a light source with a central wavelength of 1.0µm with less than 100nm width was adequate, achieving a 1.3µm longitudinal resolution.

The optical functioning principle for an OCT is a Michelson interferometer as shown in figure 1.22. This equipment analyses the interference between the light that was reflected from the retina and the identical reference light that travelled the same distance. The physics is that the light reflected by the retina is coherent, i.e. retains the optical phase that causes light rays to propagate in one or another direction, and produce an interference pattern with a coherent light travelling the same distance.

The scattered light will not produce interference and therefore will be filtered out. To produce this interference pattern, the light beam passes through a beam splitter dividing it into two directions: to the retina and to a reference mirror. The reflected beams from the retina and from the mirror are gathered in the same beam splitter, and captured by a detector. The interference pattern between the two beam lights that are in phase is their cross-correlation, as presented in figure 1.22 (*Single* reflection site).

[1] Interferometry - Property analysis technique of two or more waves by studying the pattern of interference created by their superposition.

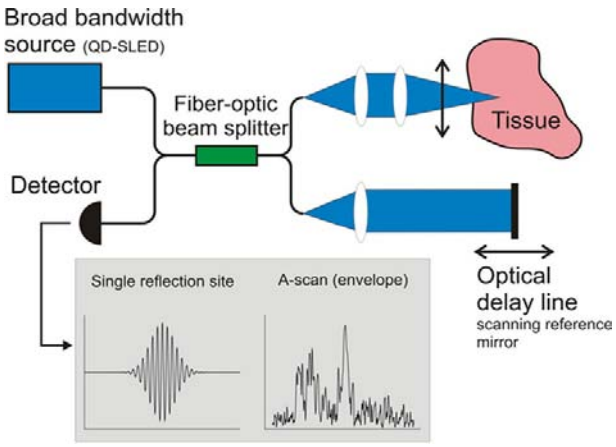


Figure 1.22 – Typical setup for an OCT. (*from* (Hogg and Andersen 2006))

By computing the signal envelope and its amplitude, it is obtained a single point of an OCT image. If the distance of the reference mirror is changed, the depth of the reflecting layer is also changed accordingly. This movement produces the so called A-scan, which is a depth scan. By moving longitudinally the sample or the optical system or capturing the interference light with a line CCD, it is obtained a cross-sectional image, called B-scan or longitudinal scan (figure 1.21). Finally the full-field OCT, which is less common, is obtained by moving in both X and Y axis generating a volumetric image (figure 1.23). Although these images have a good diagnostic potential, their visualization and comprehension are difficult, and therefore trained personnel are required.

In figure 1.21 it is shown two fundus images containing drusen in parallel with a cross-sectional OCT image. In these pictures it is very clear the presence of drusen in the layer below the retina surface. Although this equipment is very important for an accurate evaluation of drusen, by achieving micrometric resolution using an analogue, non-invasive and safe technology, the fundus photography is still the most commonly used in the regular

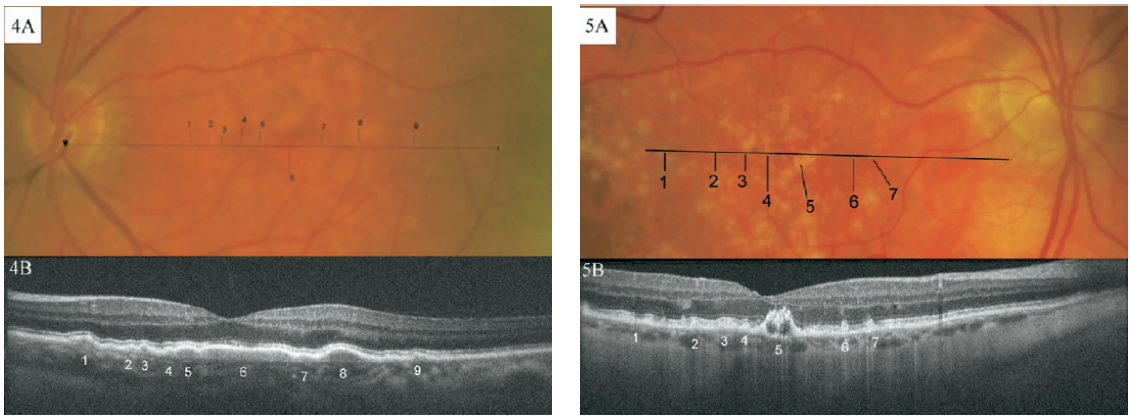


Figure 1.21 – Cross-sectional OCT images from a retina showing drusen.
(*from* (Khanifar, Koreishi et al. 2008))

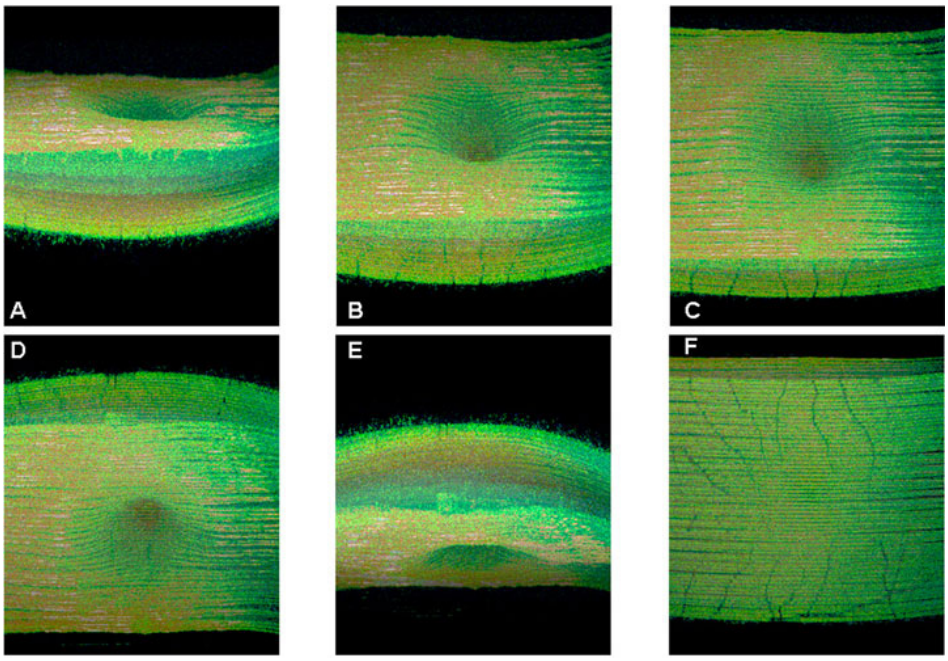


Figure 1.23 – Volumetric OCT images from a normal retina showing the fovea depression.
(*from* (Hogg and Andersen 2006))

retina examinations and on the epidemiologic studies for its wider availability.

1.3.3. Treatments

The ARMD treatments, developed so far and clinically used, are aimed to slow down drusen progression and to prevent the appearance of ARMD disease on people with high risk factors (Miller 2010). It is difficult to achieve a complete treatment, since it provokes changes in retinal structures which are not reversible. However, recent experimental studies show that damaged photoreceptors and the retinal pigment epithelium can be successively replaced using stem cells (Enzmann, Yolcu et al. 2009). Genetic studies on gene associations with ARMD (Donoso, Kim et al. 2006), which can identify the predisposal of some individuals to develop ARMD, will also play an important role as a new treatment option in the following decade.

For controlling the progression of *wet* ARMD (Choroidal Neovascularisation) there are treatments using Laser Photocoagulation that destroys new blood vessels preventing leakage. However, it creates blind spots and cannot be used in the macula due to its destructive effect. Other alternatives are drug based, such as Verteporfin (Visudyne, Novartis), Pegaptanib (Macugen, Eyetech), Bevacizumab (Avastin, Genentech) or Ranibizumab (Lucentis, Genentech), which control the new vessels growth. These latter have

been showing good results by controlling the progression of ARMD in 90% of the patients and by having improved the patients' vision in 70% of the cases.

The approved treatments of *dry* ARMD (drusen) include only drug-therapy. Although there were some clinical trials involving a technique named as Rheopheresis, which is similar to blood filtration removing substances considered as vascular risk factors, it had no FDA-approval[1]. The clinical trials were concluded with a correlation of improvements of patient's vision, although further developments were abandoned due to financial reasons. The administration of nutritional supplements containing antioxidants and zinc, such as the AREDS formula, lutein, and zeaxanthin, are the current treatments used to stabilise the progression of drusen before advancing to the wet stage, which are obtaining good results. In the US a second large multi centre study conducted by the National Eye Institute (2009) was launched to assess the effect of these nutritional supplements, as a consequence of the success of the first edition.

1.4 - The use of image processing techniques for drusen quantification

As it was shown in previous sections drusen are one of the risk factors for ARMD. As the disease has no effective treatment the only solution is to control its progression. This is one of the roles for the ophthalmologists which follow their patients with ARMD, so one can evaluate the progress of drusen. In order to perform an accurate assessment it is required to quantify the affected area and compare it along the treatments.

The quantification of drusen from fundus photographs using manual methods over fundus photographs is a choice for the ophthalmologists. The procedure is to manually draw drusen contours over fundus images and quantify the areas, but it is not commonly used as it is a fastidious process, and lacks reproducibility. Also, the variability within the analyses performed by different ophthalmologists limits the use of this technique, which will be further described in the next chapter.

There are few grading centres which are specialized in the characterization of drusen in fundus photographs using a semi-quantitative way. They classify it according to an international classification system (Bird, Bressler et al. 1995) which grades drusen in soft and hard categories and evaluates their number of occurrences. However, it is not feasible to send all the funduscopic examinations to these centres, due to delays and to the necessary logistic. As consequence, these centres are specialized on clinical trials and on epidemiological

[1] FDA-approval – America Food and Drug Administration authority that qualifies drugs and food as safe to be ingested.

studies. The common practice for the general ophthalmologist is to evaluate qualitatively the drusen progression and then act accordingly.

1.5 - What is proposed in this thesis?

This thesis proposes the use of advanced digital image processing techniques to assess the drusen affected area from fundus images. This tool will certainly help the general ophthalmologists in the follow-up of their patients, by providing them a method, which is accurate and reproducible, to measure the progression of the drusen area. It should be noticed that any technique that is developed for this purpose should be open for the clinician to validate and eventually to make adjustments to the detection suggested by the computer.

The quality of the fundus images are many times poor being affected by the optical characteristics of the retinograph and by the lack of collaboration of the patient during the exam. A misalignment between the patient eye and the flash light or a contracted pupil can produce images with non-uniform illumination patterns which can complicate their automatic analysis.

The automatic quantification of drusen that is proposed herewith is adapted to be used with fundus photographs, the most common within the ophthalmic community. The main goal was to create a method which is less parametric as possible. Parameters should be calculated based on the image information, requiring less user intervention.

The methodology includes a pre-processing step for normalization of the image intensity and contrast in order to prepare the image for further processing. A new algorithm for detection of drusen based on the labelling of gradient paths was developed, providing the drusen location and size for a modelling stage. In this latter the drusen are modelled using mathematical functions, providing an analytical characterization. By using this model the affected area is finally quantified.

This methodology was also validated clinically with the collaboration of several ophthalmologists who marked the same image set as the automated method for comparison purposes. They marked the images in the computer using a semi-automated method developed specifically for the drusen quantification. Later these images were compared to the automated method using different statistical techniques.

1.6 - Thesis structure

In chapter 2 a literature review where related works on drusen quantification and retinal image pre-processing will be covered. The methodology for drusen detection and

quantification will be described in chapter 3, being divided into five subsections. In these subsections it will be introduced the methodology steps, starting with the image pre-processing, followed by the drusen detection, drusen modelling and finally the drusen quantification methods.

The validation of the drusen automated quantification methodology will take place in chapter 4 where the prototypes are described and in chapter 5 where the validation studies and their results will be presented and discussed. The conclusions and future work will appear in chapter 6 and 7 as a closure to this thesis.

Chapter 2 -

Literature Review

The automatic detection of drusen has been object of research for almost three decades, where articles dated since 1986 can be found. The challenge is to develop methods using digital image processing that can consistently detect drusen on retinal images. These methods are not only used to detect the presence or absence of drusen, but also to determine their number, size and type. A good repeatability is also one of the pursued targets in order to maintain the same criteria among several analyses. The ability to study the drusen progression on series of images during a treatment without changing the detection criteria is very useful for ophthalmology specialists, who can accurately detect if the disease is stable or progressing.

The purpose of these researchers is to find a Gold Standard[1] Technique for the quantification of drusen. However, the subjectivity of the analysis and the need of complementary exams to support the analysis, turn it difficult to achieve a consensus among ophthalmologists of which technique is the most reliable. The areas of interest in the retina, the shape and size of drusen are some of the factors that differ among these specialists, mainly due to their different learning background (Bird, Bressler et al. 1995). The adjustments of the image slide projector or of the computer monitor and exterior light conditions are also some of the external factors that influences the detection criteria (Krupinski, Weinstein et al. 1996). These conditions should be maintained constant for every image being analysed, which is almost impossible to achieve.

[1] *Gold Standard* - denotes a practice or principle that is a model of excellence and which *objects* of the same class can be compared to.

Another aspect is that most cameras in the retinograph equipments, available in ophthalmologists' practices, are still using analogue acquisition. Some are used to capture to photographic slides and others analogue video which is displayed in a TV monitor and/or is printed on paper. Therefore, in order to apply digital image processing these images need to be digitalized first. A way to improve the use of digital image processing tools in the Ophthalmology practice is the reconversion of the analogue systems into digital image acquisition ones, in order to increase the number of image sources.

With the advances during the last decade in computing facilities, and image processing techniques, the interest in this subject has grown being published several articles on this subject, even though there are not many ongoing works on the subject (Duanggate and Uyyanonvara 2008). The ophthalmologists are requiring tools that can reproduce their analyses over their patient retinal images. Nowadays, these quantitative analysis are mainly employed on scientific studies of disease progression and on evaluation of automatic detection tools. The fastidious process of grading the retinal images is not compatible with the standard ophthalmology appointment. Therefore, to introduce these new tools into the everyday interventions, they must be user friendly, being fast at processing, and accurate for clinicians to rely on them. A chronological overview of previous works on drusen detection is presented on the following sections.

2.1 - Drusen Detection

The drusen detection methods aim to quantify the presence of the abnormalities, and additionally to characterize them according to their shape and size. Their classification into one of the two main drusen categories is also targeted by the detection methods. However their differences, contour smoothness and shape are hard to be recognized by computer software and are normally only distinguished on manual detection.

In this chapter, some of the works and techniques that were developed in the past to detect and quantify drusen will be described.

2.1.1. Manual methods

The manual detection of drusen is especially used in clinical trials for new drugs and treatments, and in epidemiologic studies to detect the prevalence of ARMD risk factors over the population. Some factors that are studied to be correlated to ARMD are the ethnic group, the unhealthy lifestyle and the presence of drusen.

Over the past twenty years there were more than fifteen studies (Schachat, Hyman et al. 1995; Vingerling, Hofman et al. 1996; Pagliarini, Moramarco et al. 1997; Delcourt, Diaz et al. 1998; Friedman, Katz et al. 1999; McCarty, Mukesh et al. 2001; Oshima, Ishibashi et al. 2001; Klein, Klein et al. 2002; Jonasson, Arnarsson et al. 2003; Varma, Fraser-Bell et al. 2004; Buch, Nielsen et al. 2005; Ferris, Davis et al. 2005; Skenduli-Bala, de Voogd et al. 2005; Augood, Vingerling et al. 2006; Wang, Rochtchina et al. 2007; Bressler, Munoz et al. 2008; Chen, Cheng et al. 2008; Hubbard, Danis et al. 2008; Xu, Li et al. 2008) all over the world with more than one thousand patients each, and the majority with more than three thousand patients (table 2.1 and figure 2.1). Considering that both right and left eyes of each patient were analysed, the size of this bank of images doubled, making their analysis an arduous task.

In these studies the maintenance of the analysis criteria was of most importance, not only to achieve accurate results, but also to allow the comparison among different studies. In most studies, the images were analysed by two different graders to guarantee the accuracy, and when their analyses differed, a third grader, usually a trained ophthalmologist, defined the final grading. Some studies with only one evaluator, applied a statistical evaluation over a selection of random images to maintain the criteria during the grading period. It was revealed in some studies the agreement among graders using the *kappa* statistics. In the Copenhagen Eye Study it was achieved an agreement of k=0.97, which is a high coefficient, although only

Table 2.1 – List of selected epidemiologic studies where drusen was evaluated on the last twenty years.

Study	Country	Date	Nr. Patients	Image[1]	Region of Interest	Drusen classification[2]	Nr. graders	Interobserver variability
The Italian eye study	Italy	1991	368	*A*	5600µm	3 levels*	2+1	k=0.68
The Barbados Eye Study	USA	1992	3444	*A*	NA	3 levels	2+1	k=0.54 - 0.80
The Salisbury Eye Evaluation Project	USA	1993	2520	*A*	6000µm	1+3 levels	2+1	k =0.72
The Visual Impairment Project	Australia	1995	4744	*A*	6000µm	3 levels	2	NA
The Reykjavik Eye Study	Iceland	1996	1021	*A*	6000µm	3 levels	2+1	NA
The POLA Study	France	1997	2196	*A*	6000µm	2 levels	1+1	NA
The Hisayama study	Japan	1998	1486	*A*	NA	3 levels	2	NA
The Baltimore Eye Study	USA	1999	5308	*A*	3000µm	3 levels	NA	NA
The Beaver Dam Eye Study	USA	1990-2000	4926-2764	*A*	6000µm	4 levels	2	NA
The Shihpai Eye Study	Taiwan	2000	1105	*A*	NA	NA	NA	NA
The Age-Related Eye Disease Study	USA	2001	4757	*A*	3000µm	3 levels	NA	NA
The Copenhagen City Eye Study	Danmark	1988-2002	946 - 359	*A*	6000µm	4 levels	1	k=0.97
The European Eye Study	NO, EE, FR, UK, IT, GR, SP	2002	5040	*D*	6000µm	3 levels	2	k=0.79
Los Angeles Latino Eye Study	USA	2003	7789	*A*	6000µm	3 levels	2	NA
The Rotterdam Eye Study	Nederland	1995-2005	6125-3761	*A*	NA	3 levels	1	NA
The Beijing Eye Study	China	2001-2006	4378-3218	*A*	6000µm	NA	NA	NA
The Age-Related Eye Disease Study 2	USA	2008	ongoing	*DA*	NA	NA	NA	NA

Legend:

[1]Image
- A - Analogue
- D - Digital
- DA - Both digital and analogue

[2]Drusen Classification
- 3 Levels - < 63 µm, 64–125 µm or > 125 µm
- 4 Levels - < 63 µm, 64–125 µm, 125-250 µm or > 250 µm
- 1+3 Levels - none, < 63 µm, 64–125 µm or >250 µm
- 2 Levels - < 125 µm or > 125 µm
- 3 Levels* - < 50 µm, 50–500 µm or > 500 µm

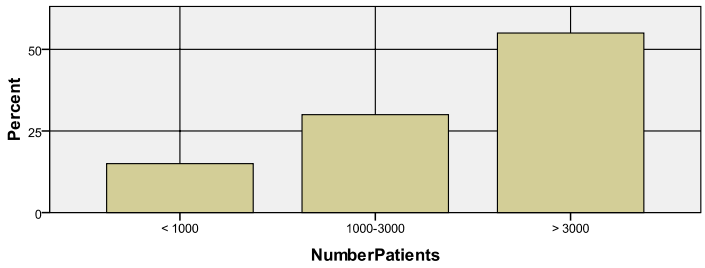


Figure 2.1 – Distribution of the number of patients in the selected studies.

one grader was involved and the number of images was less than one thousand. Other studies ranged between k=0.54 and k=0.8 depending on drusen types (smaller drusen had less agreement among the graders).

The Department of Ophthalmology and Visual Sciences of the University of Wisconsin has been one of the reference centres of the eye epidemiologic studies. They participated in studies involving a large number of patients, such as the Beaver Dam Eye Study (4926 patients) and the Age-Related Eye Disease Study (4757 patients). They also developed a protocol for fundus image analysis that was adopted by the majority of studies all over the world (Klein, Davis et al. 1991). This protocol includes a macular grid, to be centred on the macula and which defines the subfields, and a set of standard circles with different diameters to assess size and area of drusen (figure 2.2).

This division of the macula into nine subfields is clinically significant because it allows the study of relations between the location, progression and severeness of ARMD (Sarks, Sarks et al. 1994; Wang, Chappell et al. 1996). In a clinical point-of-view the central circle of the macular grid is the most important region as it contains fovea, the highest density of photoreceptors and is responsible for the detailed vision. It is followed, in terms of importance, by the superior and inferior subfields which are the ones containing the main vessels that supply blood to the fovea and, in case of containing drusen, it can be more easily propagated to the central field.

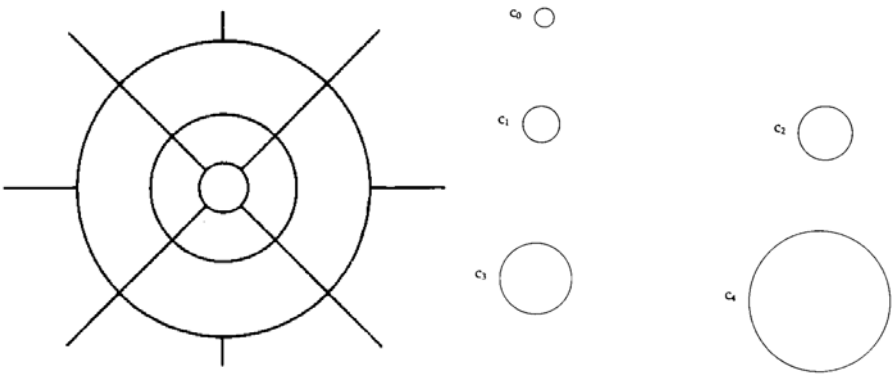


Figure 2.2 –Wisconsin grid and standard circles (from Klein *et al.* (1991)).

The Wisconsin protocol was further specified by the International ARMD Epidemiologic Study Group, which included the University of Wisconsin, leading to the International Classification and Grading System of ARMD (Bird, Bressler et al. 1995). It specifies the parameters to evaluate the severity of ARMD, such as a classification scheme for drusen, hyper/hypopigmentation, Geographic Atrophy and Neovascular ARMD. The drusen classification is based on a semi-quantitative analysis that categorizes the number of drusen (0, 1-9, 10-19, >= 20), the predominant type, the druse size (<63µm, 63-125 µm, 125-175 µm, 175-250 µm, >250 µm), the main location of drusen and the area covered on this latter (<10%, <25%, <50%, >= 50%). Using this classification scheme, the accuracy in determining the drusen area is moderate (Scholl, Peto et al. 2003), but has been sufficient for these large scale studies, even though a quantitative method is always preferable to evaluate them.

To follow up clinical studies, where progression of the total affected area is analysed, as well as to evaluate automatic detection software, a quantitative evaluation of drusen is required. The manual detection is the most common method. The fundus images are either marked in photographic paper, and then digitalized for drusen quantification (Morgan, Cooper et al. 1994) or are digitalized, and then processed using a computer assisted procedure, preferably a graphic tablet (Smith, Chan et al. 2005).

The specialist that uses photographic paper, draws the drusen inside contour in a clear transparency placed over the printed image (figure 2.3). The transparency is afterwards digitalized and requires basic image processing to fill the contours, to segment the image and to quantify the areas. This image preparation and area quantification must be assisted by an experienced user to guarantee the final quality. It is a reliable and practical procedure, although any retouch to the contours is difficult to execute and requires the supervision of an image processing expert.

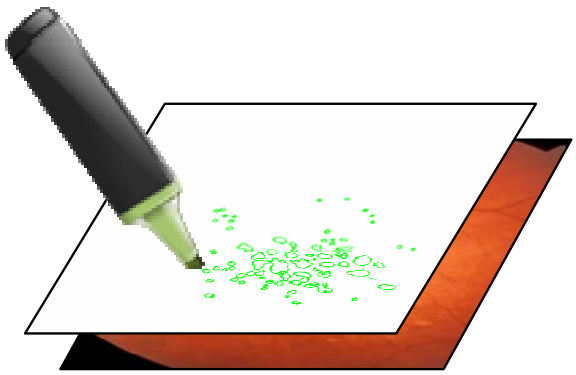


Figure 2.3 – Manual drawing over printed fundus photographs.

The second procedure uses computerized tools to draw drusen contour in a manual or semi-automatic method. If the fundus photographs are not in digital format, they are either digitalized from paper or from 35mm slides. Following, a druse contour is drawn on an image editing software, followed by its inside fill up. Once all drusen are detected the painted pixels are accounted to obtain its total area. With the increasing availability of computerized tools, which are more flexible and productive, this latter method is the most commonly used within the manual methods.

Unfortunately, these manual techniques are influenced by the subjectivity of operator's criteria. This fact increases the variability among the ophthalmologists, which limits the accuracy and reproducibility of the measurements. However, these are considered the gold standard techniques for the drusen detection and analysis.

2.1.2. Automatic detection Methods

The automatic detection methods aim to recognize drusen in retinal images and to quantify the affected area using image processing tools without any user intervention. The complexity of the retinal images associated with the illumination non-uniformity, which affects the majority of the images, along with the different pathologies that are visible on the retinal surface, make the development of a computerized method that can distinguish drusen a difficult task.

Although the research focus is the development of unsupervised methods, the user supervision for validation and results adjustment purposes should not constitute a rejecting factor. Actually, to get a better acceptance from the clinicians a method with minimal supervision is recommended. With other diagnostic techniques the clinicians can correlate both results and validate the automatic detection.

The first important work in detection of drusen was published in mid-80's by Peli and Lahav (1986), being a reference to subsequent works. The drusen detection techniques are mainly threshold based methods, some other with a post-processing to reject non-drusen spots. In the following subsections a survey on automatic drusen detection methods will be presented.

2.1.2.1 Automatic Thresholding

The use of image threshold techniques has been the most common approach to the drusen detection problem. The contrast between the drusen and the background is apparently adequate for the threshold technique. However, the contrast and illumination is not constant

throughout the image what invalidates the use of an absolute image threshold. The strategy most widely adopted is the application of the thresholds values that results from the local illumination and contrast evaluation.

The first important proposal that use this technique was published by Peli and Lahav (1986) and further developed by Sebag *et al.* (1991) from New England Medical Center and Tufts University School of Medicine. The authors presented an adaptive thresholding method that divided the image in non-overlapping 8x8 pixel windows, and evaluated its histogram distribution looking for drusen indicators. It was assumed that a window containing both background and part of a druse has a bimodal distribution, and that it could be detected by analysing its histogram standard deviation. In these cases the threshold value is the median of the grey levels. The detection of false positive or false negative bimodal histograms was considerable, due to the presence of vessels or small parts of drusen. These false detections had to be corrected by the operator in each of the incorrect 8x8 pixel window where it has occurred. After all bimodal windows are identified, corrected by the operator and computed their thresholds, these are smoothed by a two-dimensional linear interpolation. The smoothed threshold values were then applied to each individual window.

The authors evaluated the reproducibility of their technique on three repeated images of seventeen eyes using the *coefficient of variation* (2.1). A reproducibility of $R = 4.7 \pm 3.5\%$ was obtained, which showed the reliability of the technique. The user supervision and corrections were always needed, limiting the use of this method on large datasets.

$$coefficient\ of\ variation = \frac{standard\ deviation}{mean} \times 100\% \qquad (2.1)$$

A multi-disciplinary research group from the University of Aberdeen, which included the Department of Ophthalmology and the Department of Bio-Physics, demonstrated (Phillips, Forrester et al. 1993) that a combination of global and local thresholds applied in retinal images can be used to detect drusen. With the green channel they apply a shade correction algorithm to correct non-uniformities of illumination, obtaining a flattened image (Jagoe, Blauth et al. 1990). Large confluent of drusen are then detected by thresholding the image at an intensity of (*mode + 50*). For the smaller and lower intensity exudates' detection, the region of interest is divided into 32x32 pixel windows which are thresheld at a value of (*mode+20*). The two thresheld images are combined to represent the areas affected by drusen.

The results obtained were compared to the gradings made manually using a transparency on a set of 13 colour slides. As for reproducibility tests, two slides four days apart were taken reflecting a sensitivity average of 87% and a reproducibility (computed using

equation (2.1)) range of 17% for small areas and 5% for large areas. One of the disadvantages found on this approach is the discontinuity provoked by the use of threshold values, independent for each window. Additionally, a cause for false positive detections is the influence of vessels branching and crossing, which lowers the threshold reference value and consequently detects false drusen.

Later on, Morgan *et al.* (1994) proposed an automatic drusen quantification method based on the application of the Otsu threshold method (Otsu 1975) to local windows of 16x16 pixels. The images were grey scale digitalized slides, illuminated by red-free light to improve the contrast. The procedure presumes that the histogram of an image with vessels, background and drusen contains three lobes (one for each) separated with two dips. It was then divided into two steps: vessels detection and drusen detection. To find the threshold values that separate these three lobes, the image was divided into 16x16 pixel windows and for each pixel it was determined the threshold value based on the block's Otsu threshold as well as of the adjacent blocks values, weighed by the pixel's location. With this procedure the authors could smooth the discontinuities that were usually observed with local windows' threshold.

The image that lies below the obtained threshold map contains theoretically the vessels. Hence and to detect drusen, the threshold determination is performed again on the image that lies above the threshold map, which is the background and drusen. Finally, this threshold map is elevated by 16 grey levels to achieve a compromise between false-positive and false-negative detection and is applied to generate a binary image where drusen are detected.

The results were only compared to specialists' analyses on drusen counting. Although, it detected the same drusen as the specialists, the computer procedure had predominantly detected more drusen and consequently the agreement was low (Friedman χ^2=10.54; *df*=3; P=0.05). Further analyses were missing to evaluate the sensitivity and specificity.

A work published in 2003, carried out at University of Crete by Rapantzikos *et al.* (2003), uses image enhancement techniques to compensate illumination and to evince the retinal structures as well as a combination of global and local thresholds to detect drusen. Following the image pre-processing step, a two stage threshold which starts with a global threshold (Otsu) and then with a histogram adaptive local threshold (HALT) is applied. This adaptive threshold inspired by the work of Shin *et al.* (1999) divides the image into nine sub-

windows and computes their local histogram. Besides the histogram's *skewness*[1], already used in the previous work, the authors also used the histogram's *kurtosis*[2] to evaluate the presence of drusen in the sub-windows and chose the according threshold method. Between the global and the HALT thresholds a morphological dilation is applied in order to expand regions that were not removed by the global threshold.

The algorithm was tested in a set of twenty three images and compared to two specialists' analyses. For the specificity and sensitivity analysis the interception between the specialists' gradings was used. This decreased the probability of having false negatives and consequently a higher sensitivity. These two indicators were evaluated and exceeded 96% in all cases, which can be considered excellent results. Although, it should be noted that the specialists' interception is not a reliable source, since it eliminates the variability and increases the probability of higher sensitivities without compromising the specificity.

Smith *et al.* (2005) proposed an iterative algorithm to detect drusen by levelling the background and extracting drusen. Their method divides the image into two zones: the central-subfield (zone 1), which does not contain vessels and the annulus of inner and outer diameters of 1000 and 3000μm (zone 2), which does contain vessels. On zone 1 they apply the Otsu method with two classes C_1 and C_2 (one threshold value) to separate background (C_1) from drusen (C_2). As for zone 2, they apply the same method with three classes C_0, C_1 and C_2 (two threshold values) to separate vessels (C_0), from background (C_1) and drusen (C_2). Then the image containing only background pixels is modelled by a 2D quadratic function using a least-square fitting method. In order to correct the non-uniform background, the generated model is subtracted from the original image. This process of thresholding, modelling and subtraction is repeated until the background model is stable between iterations. Finally, the drusen are detected by the Otsu threshold. For some images, with few drusen, the Otsu method is not able to find the drusen-to-background threshold and adds background and drusen to the same class (C_2). To such cases, the authors propose a manual intervention specifically on class C_2, where the Otsu method with a single threshold value to separate drusen from background is reapplied.

The method was tested with a set of twenty images, which were also manually graded by two specialists in conjunction to achieve a consensus. The algorithm's sensitivity and

[1] *Skewness* - is a measure of the asymmetry of a probability distribution. (*negative skew* – the left tail is longer; *positive skew* – the right tail is longer)
[2] *Kurtosis* – is a shape factor measure of a probability distribution.(*mesokurtic* – normal distribution, *leptokurtic*-long tails and narrow peak, *platykurtic*-short tails and broad peak)

specificity were computed and obtained 0.70 and 0.81 median values, respectively. The low specificity indicates that the algorithm over-detects drusen. The Otsu method is better suited to images containing intensity histograms with fixed number of groups, which is not the case of retinal images. With these the method can generate erroneous thresholds that require user's supervision to produce reliable results.

2.1.2.2 Region growing

Kirkpatrick *et al.* (1995) presented a technique to quantify drusen on SLO images and compared it to the technique previously presented by Philips, Forrester *et al.* (1993) which quantified drusen from fundus photographs. For this comparison the authors took fundus photographs and SLO images of six patients on the same day.

The retinal indirect SLO images show a *«"moon surface" appearance, with drusen edges appearing as "shadows"»* (figure 1.19.b) and present a more uniform illumination than fundus photographs. The authors proposed an algorithm which uses gradient edge detection for drusen boundaries and a region growing method to fill their centres. The region growing method is divided into four steps, each one for each side of the druse (top, left, bottom, right). Each step starts with a binary image containing only one side of the druse as the method seed. Then it grows to the opposite side until the gradient exceeds a predefined threshold.

The results showed that the accuracy was lower than acceptable, as at a specificity of 90% the sensitivity was only 35% for the SLO method, in contrast with the fundus photographs method that achieved 60%. However, these results were not statistically significant as only six images were analysed. Several reasons for the low sensitivity were raised; however the main cause was the inappropriate edge detection method applied to find drusen contours, namely soft drusen.

An adaptive local threshold that considers the local histogram *skewness* for detecting the presence of pixels belonging to drusen complemented with a region growing algorithm was presented by Shin et al. (1999) from University of Pennsylvania, School of Medicine.

The algorithm uses an image pre-processing based on noise reduction filters and background illumination compensation before the drusen segmentation. For the segmentation the region of interest (a circle with 1500µm radius centred on the macula) is divided into smaller areas, which vary in size from 20x20 pixels to 100x100pixels where the areas' histograms *skewness* are evaluated. The authors verified that the areas containing drusen, that are brighter than the background, have a *skewness* higher than 0.5. In these latter they apply a

local threshold using a value based on the mean intensity of the region of interest plus a term proportional to the areas standard deviation.

To improve the algorithm sensitivity an intensity-based region growing is applied to the detected areas to expand the incompletely identified drusen. The neighbour pixels that fall within a similar threshold (which can be defined by the grader) are identified as drusen. At the end the grader can adjust the results by adding or removing pixels, incorrectly identified.

To test the algorithm, four photos, two for each eye, separated by one year were taken from five patients. From each pair, only the eye that the authors considered to be the best captured was selected, making a total of ten images. The results from the manual grading, done by one specialist, were compared to the results produced by the algorithm when *unsupervised* and when *supervised* by two specialists. The agreement was evaluated by the *Intraclass Correlation Coefficient (Fisher 1970)*, which obtained 0.67, 0.93 and 0.92, respectively for the *unsupervised* version and *supervised* by each specialist. These values translated a high consistency when *supervised*, but medium performance when *unsupervised*.

Soliz *et al.* (2002) from University of Pennsylvania presented a region growing segmentation algorithm for drusen quantification. It included an illumination and acquisition artefacts correction, by applying a Gaussian low-pass filter on the green channel, and by analysing the blue channel, respectively. The region growing seeds were found when dividing the image according to the Wisconsin Grading System grid regions, and choosing a threshold value based on a statistical analysis of the local histogram.

The region growing method is based on the Savol's adaptive object growing algorithm (Savol, Li et al. 1980). It uses the highest maximum intensity pixel, previously detected, as first seed. In case its adjacent pixels perform an increase or maintenance of the cluster's overall contrast, then these same adjacent pixels will have the required characteristics to be added to the cluster. At the end of this process a new calculation restarts on the next cluster.

Their algorithm was applied to a set of eighteen images that were then supervised by an ophthalmic specialist to improve the segmentation results. Both the results before and after supervision were compared to the ones produced manually by a specialist that followed the Wisconsin Grading System, i.e., classified drusen according with the largest drusen found on each of the grid regions. The *kappa* statistics was used to measure the agreement and disagreement between them. The overall agreement with the manual grading was higher with user supervision, achieving a maximum weighed agreement of 0.712. Although, the technique presented by the authors obtained good results, a more accurate comparison of the detected areas was missing.

2.1.2.3 Fuzzy-logic thresholding

A fuzzy logic approach to the problem of drusen detection was presented by Thdibaoui *et al.* (2000) from the University of Paris. The subjectivity of drusen analysis motivated the authors to apply fuzzy logic techniques to classify the pixels. The algorithm is divided into two stages: pre-segmentation and fuzzy segmentation.

The initial stage is to find an optimal partitioning for the image histogram into three intervals, using a minimization of mean square error strategy. The first interval (darker pixels) stands for background pixels, the last interval (brighter pixels) represent the pixels belonging to drusen and the intermediate interval to the ambiguous pixels. A fuzzy-logic approach will be used to classify these latter accordingly. The final degree of belongingness of each pixel is calculated considering the mean degree of belongingness in the pixel's neighbourhood of k pixels. Unfortunately, the presented work missed a quantitative evaluation of the algorithm and therefore was not scientifically validated.

2.1.2.4 Morphologic Geodesic Reconstruction

A geodesic reconstruction for drusen segmentation was proposed by Sbeh *et al.* from University of Paris-Dauphine (1997; 2001). The algorithm, after applying a pre-processing step for image smoothing, detects the image regional maximums. These are then used as seeds for a geodesic dilation procedure, which include (*x*,*y*) expansion and contrast increase, and stops when the dilated pixels intensity are below the original image intensity. From the generated synthetic image (*geodesic neighbours*) the background is extracted and the drusen detected. Their algorithm finishes with a post-processing to eliminate those objects that are not related with drusen. Their classification criterion includes the object's *shape*, minimum *contrast* and minimum *area*.

The overall results were considered good; nevertheless no quantitative comparison with manual techniques was presented.

2.1.2.5 Wavelets

A Multi-level Analysis which included wavelet processing was proposed by Brandon and Hoover from Clemson University (STARE 2000; Brandon 2003; Brandon and Hoover 2003). Their multi-level algorithm performs a classification which begins at the *pixel* level and then proceeds to the *region* level, *area* level, and finally *image* level. The *pixel*, *region* and *area* levels are intended to detect and filter pixels/areas which correspond to drusen,

while in the *image* level it is given an overall image classification on the amount and type of drusen present in the image.

As a pre-processing step, they used the green channel and applied an intensity based image equalization and a wavelet-based equalization. Further details on this pre-processing can be found in Hoover and Goldbaum article (2003). The *pixel* level detection of drusen is done by applying the wavelet processing. The idea is that a pixel within a druse should exhibit a characteristic response to a wavelet transform. The *region* level analysis groups the output pixels from the previous step to form regions. These regions are characterized with measures such as area, average intensity, maximum intensity, borders intensity, average gradient, border gradient and others in order to qualify it as a druse. Non-qualifying regions are removed from the output image. The *area* level analysis is similar to the previous one except it analyses and filters the regions inside a window of 100x100 pixels centred on each region and classifies the regions in one of the four categories: *large many*, *fine many*, *large few* and *fine few*. As for the *image* level, it compiles the regions information and classifies the image based on the number of regions of each type in one or more categories: *normal*, *large many*, *fine many*, *large few* and *fine few*, *abnormal*.

The algorithm was tested in one hundred and nineteen images from which sixty two contained drusen, thirty eight were healthy retinas, and nineteen contained non-drusen abnormalities. The results were compared to a ground truth provided by one ophthalmologist, which classified the images in one of the *image* level categories. Hence in their presentation of results only the *image* level classification was evaluated. The results were analyzed with two measures *Drusen vs. Non-Drusen* and *Exact classification*, obtaining 87% and 71% of agreement, respectively. Unfortunately, these results are not interesting to evaluate the drusen detection and quantification method due to its lack of specificity.

2.1.3. Considerations

This presentation of researches on drusen detection showed that this is a difficult problem to solve. The majority of the works are based on image thresholding techniques and only some are based on region growing algorithms or morphological operators. Also, the post-processing that removes incorrect detected drusen to increase the algorithms' sensitivity is not applied.

The algorithms' validation is not being emphasized. The methods that were so far developed presented interesting approaches, but unfortunately they were not tested with large volumes of retinal images in order to evaluate its' performance, statistically.

2.2 - Retinal Image Pre-processing

As presented in the previous section, an image pre-processing is a determinant factor to achieve a reliable detection and quantification of drusen. It can include the colour space selection and/or transformations to improve the image contrast and to show some of the retina structures. Also, the non-uniformity of illumination is one of the main factors that influence the quality of drusen detection algorithms which requires special attention. In this thesis this is also a key factor. Therefore, some of the main techniques used in retinal image pre-processing will be presented next.

2.2.1. Image contrast enhancement

The fundus images are usually acquired in colour mode and in grey scale mode on some older systems. In the colour mode, the Red Green Blue (*RGB*) colour space is the most common due to its wider availability within the cameras and the frame grabbers. However, other colour spaces, such as, Intensity Hue Saturation (*IHS*) or Luminance and Chrominance (*YCrCb*) can be used. The choice of colour mode channel(s) is important, as this can provide better quality images for automatic processing.

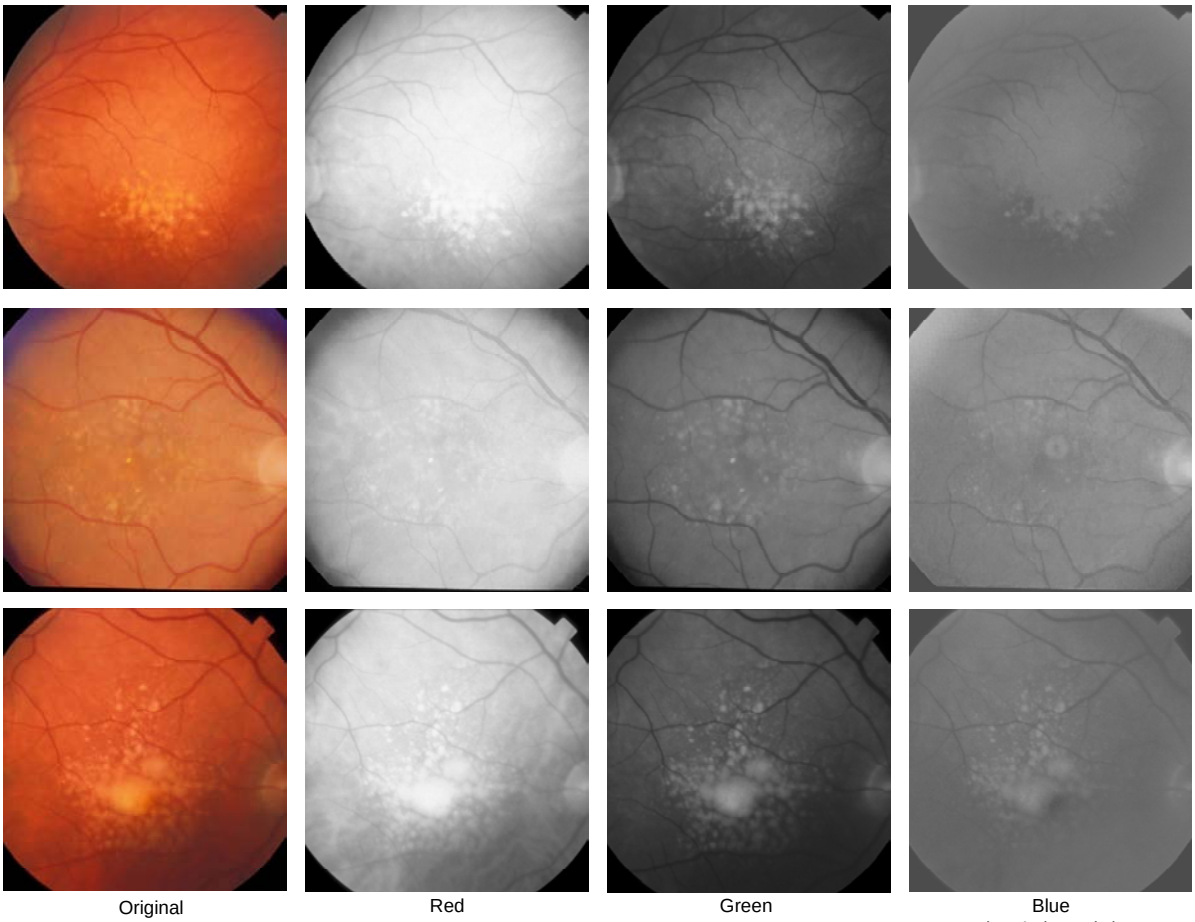


Figure 2.4 – Retinal images decomposition in RGB channels.

In spite of the reddish appearance of the retinal images in the *RGB* colour space, the Red channel is not well suited for automatic image processing, as illustrated in figure 2.4. Besides the high intensity saturation, the non-uniformity of the illumination stands out more in the Red channel than in the others. The Blue channel also contains the same influence of the non-uniformity of the illumination and is low contrasted. The Green channel, which is used in the majority of the research works (Shin, Javornik et al. 1999; Soliz, Wilson et al. 2002; Rapantzikos, Zervakis et al. 2003), has the better dynamic range, better contrast on drusen, and is less influenced by the illumination differences. This effect has been justified by a combination of several factors: the retinal structures have a higher reflectance to red wavelength than green and blue wavelengths; red wavelength enters deeper in the retinal layers; and the deeper layers are more angle-dependent which generates scattered light (Schweitzer, Schrödel et al. 1985; Delori and Pflibsen 1989).

The IHS colour space, on the other hand, does not have these contrasting differences (figure 2.5). For one, the Intensity channel is similar to the grey scale image, where the non-uniformity of the illumination and different contrasts are clearly visible. Second, the

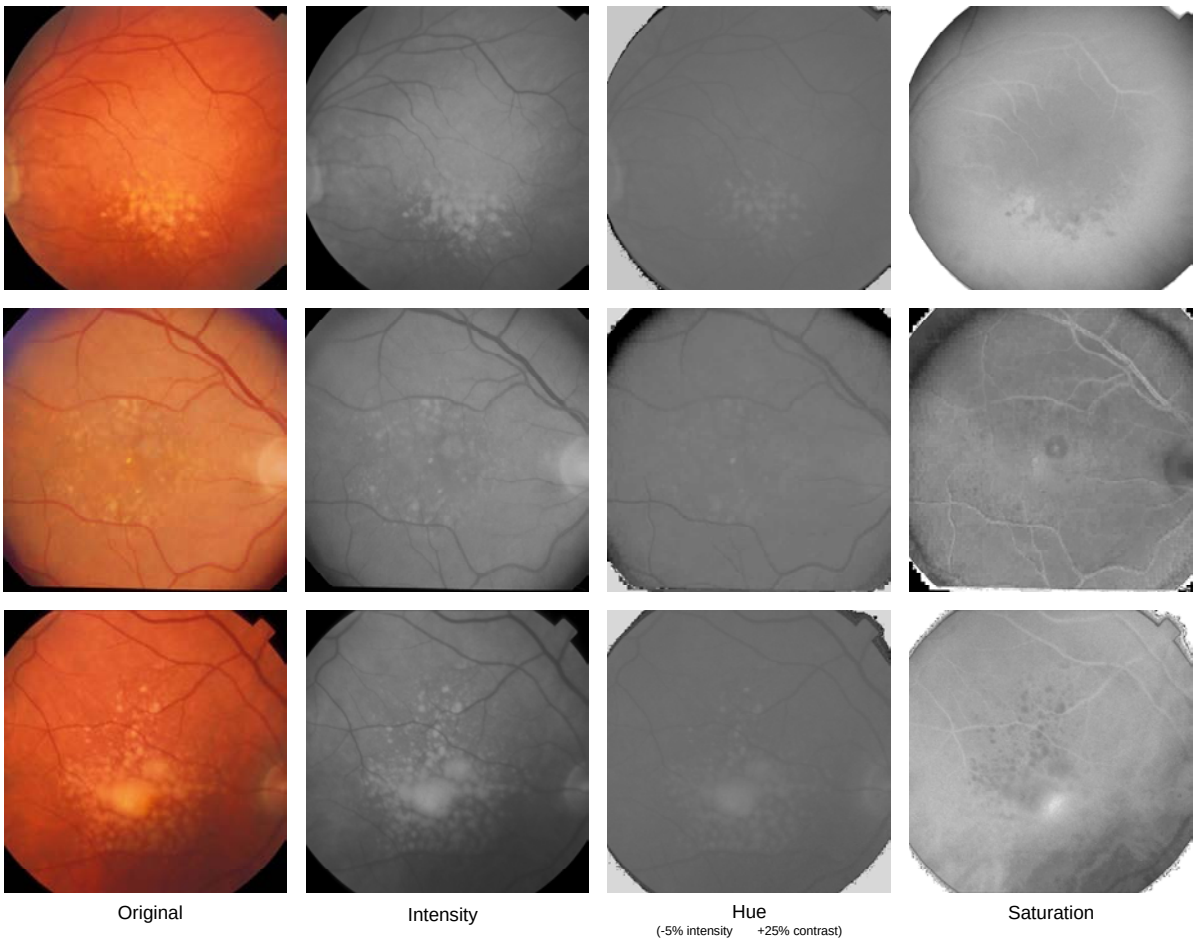
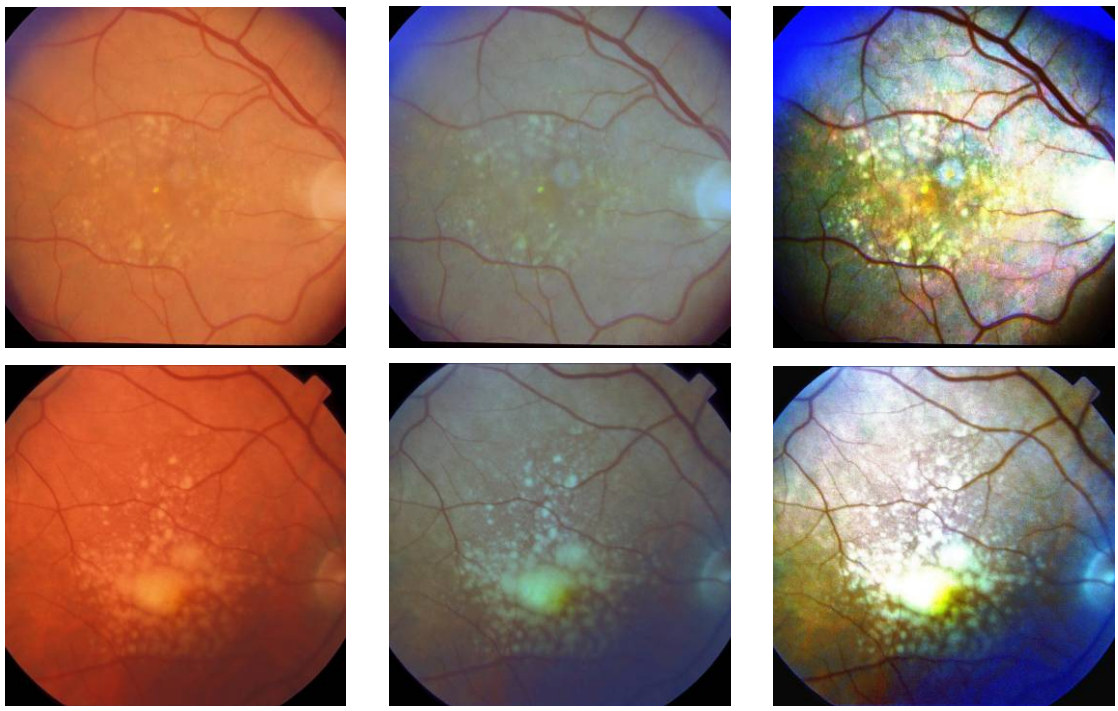


Figure 2.5 – Retinal images decomposition in IHS channels.

Saturation channel does not have useful information to the detection of any retina structure. And finally, the Hue channel is perhaps the one that has a better contrast between drusen and background, and where the illumination is uniform. Sinthanayothin, Kongbunkiat *et al.* (1999; 2003) use the IHS to enhance the image contrast, by applying a histogram equalization in the intensity channel, before the RGB image regeneration.

The image histogram manipulation is another method used to enhance the retinal images' contrast. Goatman, Whitwam *et al.* (2003) compared three different techniques which were adapted to retinal image processing. One of such techniques is the *Grey-World Transformation* (Finlayson, Schiele et al. 1998). It is based on the assumption that an image with sufficient colour variation reflects an intensity average of each *RGB* channel that should tend to the intensity average of the Intensity channel image. Thus, the transformation normalizes the *RGB* channels, by dividing each of the channels by its intensity average and by multiplying them with the intensity average of the Intensity channel image. The output is presented in figure 2.6 (a,b). This technique improves the perception of the human eye by expanding the colour range. However, improvements in the correction of non-uniformities in background are not noticed.

Another technique that was also presented in their article was *histogram equalization*, where each individual colour channel is equalized. This expands the interval between colours



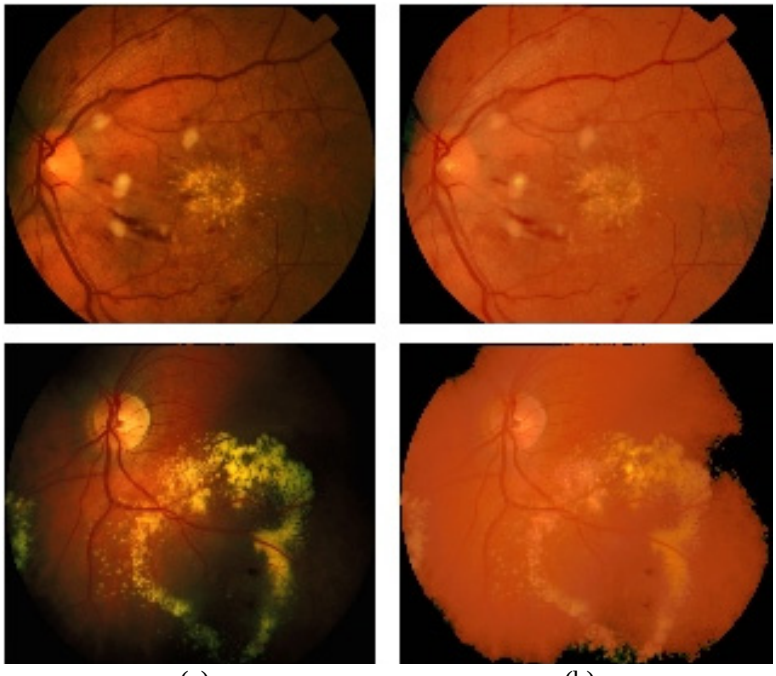
(a) *(b)* *(c)*

Figure 2.6 – Colour normalization on retinal images. (a) original image; (b) greyworld transformation; (c) histogram equalization.

increasing the overall contrast, but also increasing the severity of non-uniform illumination. Another consequence is that the resulting image is more saturated. The use of this technique on retinal images is presented in figure 2.6 (a,c). A multilevel image enhancement that combines a global and a local histogram equalization was used by Rapantzikos, Zervakis *et al*. (2003). In their algorithm, the global equalization enhances brighter parts, while the local use in smaller windows improves the contrast on the darker areas.

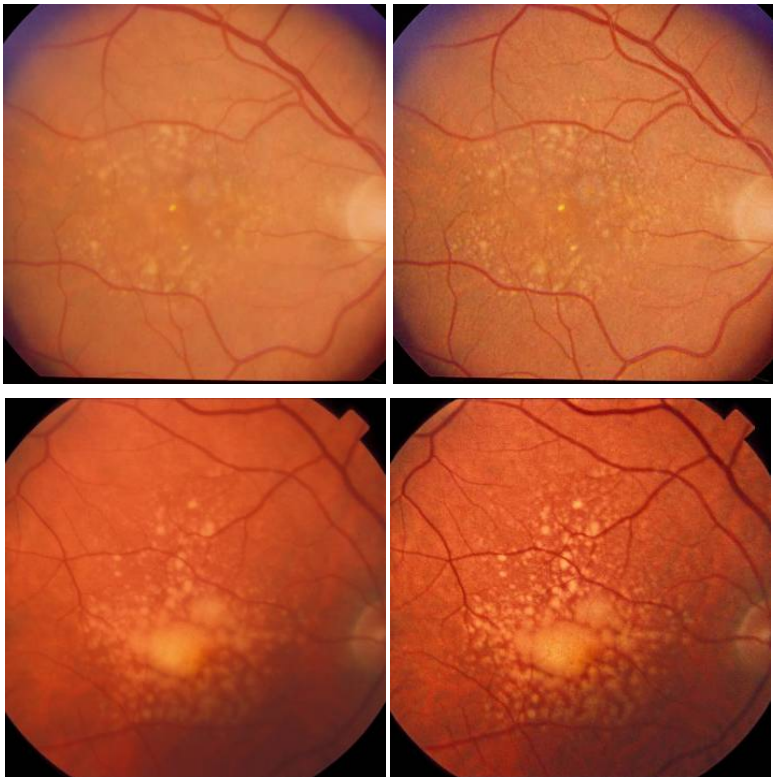
The last technique presented was *histogram specification* (Gonzalez and Woods 2007), which was used in Osareh *et al.* work (2002) for retinal exudates detection. In this technique the RGB histograms distribution are adjusted to exhibit a shape similar to reference distributions. In order to achieve more realistic images with higher contrast, the reference distributions will be the ones shown in images containing a normal colour distribution and contrast. With this technique it is possible to give different weights to the RGB channels, especially to reduce the influence of the Blue channel which in retinal images has little drusen information. The use of this technique on retinal images is shown in figure 2.7. The improvements in image contrast are noticeable and additionally it produces an image with normalized contrast which constitutes a good approach to a standard analysis.

The *top-hat* and *bottom-hat* transforms are morphological operations that improve image contrast and are used in retinal imaging, especially to show vessels and drusen (Leandro, Cesar et al. 2001; Zhang and Fan 2006). On the one hand, the *top-hat* transform is



(a) (b)

Figure 2.7 – Histogram Specification for colour normalization. (a) original image (b) histogram specification. (*extracted from (Goatman, Whitwam et al. 2003)*)



(a) *(b)*

Figure 2.8 – Morphological transformation. (a) original image (b) contrast enhanced image.

obtained by applying to the image a morphological *opening* operation, which is then subtracted to the original image. On the other hand, the *bottom-hat* transform is obtained by subtracting the original image and its morphological *closing*. These two transformations are used for image enhancement, by adding the *top-hat* image and subtracting the *bottom-hat* image to the original one.

In figure 2.8 the morphologic transformation was applied on two images. The result is similar to an edge enhancement operation where the contrast on the structure edges is increased. Although, an increase of low amplitude noise is noticed on the flatter areas and the non-uniform contrast remains unchanged, it is an operation that is useful for a manual grading, even though with few improvements for an automatic image processing.

2.2.2. Non-uniform illumination compensation

The non-uniformity of illumination is very common in retinal images. It can result from the different reflecting characteristics of the retinal surface, from problems of the eye lens, from the equipment vignetting and from poor patient collaboration in the image capture. The patient's voluntary and involuntary eye movements during the image capture can cause misalignment and interfere on the image quality. The consequences are different illumination and contrast areas over the image, which make it difficult for image interpretation and

Figure 2.9 – Retina image with non-uniform illumination. Notice the lower reflectance of the macula and the non-uniform illumination on the right side.

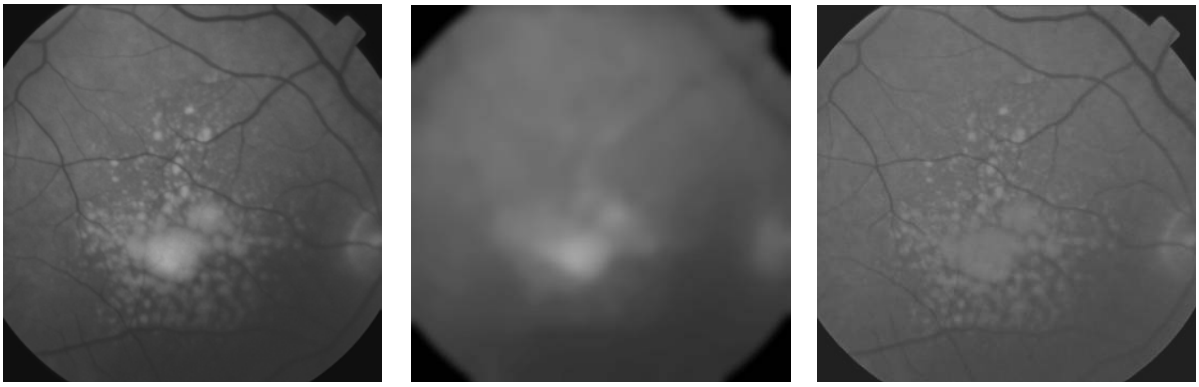
processing. Besides having brighter and darker areas, the contrast in these areas is directly related to the quantity of illumination to what they are exposed.

As these facts influence the quality of automatic drusen detection algorithms, they are usually corrected in pre-processing step before the detection algorithms. The major difficulty in non-uniform correction algorithms is to separate illumination from the reflectance. The illumination is usually a low frequency image component that can be a constant value when illumination is uniform. The illumination distribution can have complex patterns which can be non-uniformly distributed along the whole image. For example, an image can contain a uniform illumination and in one extreme have a darker area due to a camera misalignment.

However, not all the non-uniformities are random; it is known that the macula has a lower reflectance than the other parts of the retina (figure 2.9). This is due to the higher concentration of photoreceptors in this region, which have a higher absorption of visible light. It is known as the macula lutea[1], a region highly pigmented by yellow colour.

In the work of Smith *et al.* (2005) the authors studied the reflecting properties of the macula and described the lower reflectance with a series of multiplicative factors gradually decreasing when they move away from the centre. Then they applied to the Green channel these multiplicative factors, a method which was entitled as *luteal pigment correction*. Analysing their results, it can be seen that in the images where the macula is not affected by drusen, the method has a good performance. Unfortunately, drusen do not absorb the light as the photoreceptors and therefore when they are present in the macula the results of this method are limited. Also, this approach requires the user to define the exact location of the macula to be effective.

[1] Macula lutea is a yellow spot near the centre of the retina. From Latin *macula*, "spot" and *lutea*, "yellow".



(a) *(b)* *(c)*

Figure 2.10 – Correction of the non-uniformity of the illumination using Gaussian convolution.
(a) Green channel image; (b) illumination pattern; (c) corrected image (*green channel*).

Another approach to correct illumination is the image low-pass filtering. The major problem of this technique is to define the filter's cut-off frequency, in order to remove the non-uniformity of the illumination and maintain the image details. A low cut-off frequency will not correct the image; while a higher frequency can smooth the image and remove some of the retina structures that are undesirable. This is especially noticed on images with a high amount of drusen or other retina abnormalities, which influence the overall intensity.

The mostly used technique is a convolution filtering with a Gaussian kernel to estimate the image's illumination pattern and to normalize it accordingly (Jagoe, Blauth et al. 1990; Phillips, Spencer et al. 1991; Soliz, Wilson et al. 2002). The kernel used varied in size between these works but the criterion was to have a compromise between being larger than the abnormalities and smaller enough to detect local variations. The kernel values were generated using the equation (2.2), in which σ is related to the frequency interval and is determined empirically. The output from the convolution is a smoothed image without any details, which is then divided by the original image in order to normalize the illumination and the contrast.

$$G(x,y) = \exp\left(-\frac{x^2+y^2}{2\cdot\sigma^2}\right) \qquad (2.2)$$

Figure 2.10 shows the use of this technique on an image containing drusen in an advanced stage. The technique used was as described by Jagoe, Blauth *et al.* (1990). The image was reduced to one quarter of the original size, filtered by a 9x9 pixel median, enlarged to the original size, filtered by a Gaussian kernel of 15x15 pixel and divided by the Green channel. As it can be seen in figure 2.10.b, although the non-uniform illumination pattern has been detected it was influenced by the presence of large drusen.

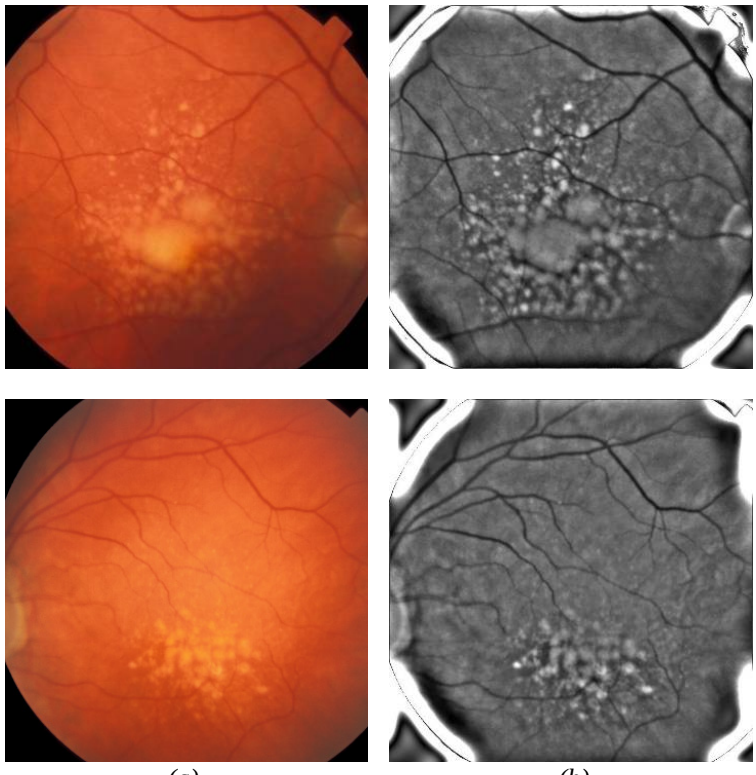
Homomorphic filtering is also a technique used to correct non-uniform illumination in retinal images (Rapantzikos, Zervakis et al. 2003). It belongs to the family of frequency filtering techniques and it is usually applied on signals composed by multiplicative factors. In the image processing context an image is roughly the product of illumination and reflectance (not considering the incidence angle), and therefore complies with the technique requirements.

$$f(x,y) = i(x,y) \cdot r(x,y) \tag{2.3}$$

Based on this assumption, *homomorphic filtering* separates illumination and reflectance and corrects the illumination accordingly (Gonzalez and Woods 2007). The technique converts the image to the logarithmic domain where the multiplicative factors are converted into additive ones.

$$\begin{aligned} z(x,y) &= \ln\big(f(x,y)\big) \\ &= \ln\big(i(x,y)\big) + \ln\big(r(r,y)\big) \end{aligned} \tag{2.4}$$

Following is a high-pass filter in the Fourier domain that amplifies the image reflectance (the higher frequencies) and attenuates the non-uniform illumination (the lower frequencies). It finishes by converting the image to the spatial domain once again, where it presents a higher contrasted image with levelled illumination.



(a) *(b)*

Figure 2.11 – Homomorphing filtering. (a) Original images; (b) Butterworth filtered images.

In figure 2.11 homomorphing filtering is applied on two images containing non-uniform illumination. The filtering was made in the Fourier domain with a third-order high-pass Butterworth filter (*radius* = 8; γ_L = 0.09; γ_H = 1.6). The parameters were determined empirically to generate a contrasted image with a levelled illumination. However, the ringing produced by the FFT filtering on the high intensity transitions is significant. If the image is centred on the macula this may not be a problem, because it does not intercept the region of interest.

It is a technique which has a good levelling capability; however a parameterization for all images is difficult. The cut-off frequency determination and the ringing provoked by the FFT are its main disadvantages.

2.3 - Summary

This chapter covered the most important works on drusen detection and on image pre-processing for fundus images. It has also analysed some of the epidemiological studies in which the drusen were graded manually and where the use of automatic drusen detection could play an important role to automate the image grading process. It was stated that image processing techniques for automating clinical analysis are still insignificant.

A chronological presentation of researches on drusen detection showed that the majority are based on image thresholding techniques. These varied from local use on smaller windows within the image to global thresholds applied to the whole image or in a region of interest, or a combination of both. Fuzzy-logic was also employed to classify confounding pixels as drusen or background after a simpler image thresholding. The majority of these works, except Sbeh, Cohen *et al.* work (2001), are not performing any post-processing to remove incorrectly detected drusen, which is important to increase the algorithms' sensitivity.

Nevertheless, the use of other techniques to detect drusen rather than thresholding is still limited to one or two works, through region growing algorithms or morphological operators. It is still an open research field, since a gold standard method for drusen detection is still to be achieved. The validation with large volumes of retinal images in order to evaluate its' performance statistically is also still limited. The lack of reference databases is considered to be the major restricting factor.

It was also noticed that a correct image pre-processing is a key factor for the method's reliability. The improvement of the image contrast and the correction of the illumination non-uniformity are the main techniques applied by the authors. These steps normalize the retinal

images for further processing, simplifying the application of the subsequent methods. For its improved drusen contrast and less dependence on non-uniform illumination, the green channel was the choice within the authors to be used as basis of the image processing. The illumination correction is an accessory method that is perhaps the one that has more improvement in the quality of the drusen detection. These techniques showed good results, although some required more difficult parameterization.

This thesis took a different approach. It is dedicated to evaluate the use of surface modelling techniques to characterize drusen. It involves the correction of the non-uniformity of illumination using smoothing splines fitting, the detection of the location of drusen spots using a newly developed algorithm, and finally the modelling of the detected spots. This allowed a more robust characterization of drusen. For the evaluation of the technique, accuracy and reproducibility were considered as the main criteria, as it will be presented on the following chapters.

Chapter 3 -

The Methods of Detection

The aim of this work is to provide ophthalmologists a tool that quantifies drusen in retinal images, which they can rely on to maintain the same analysis criterion for a patient's sequence of images. It should be noticed that more important than knowing the exact area affected by drusen, is to guarantee that a uniform criterion is used in every image despite the clinician who is applying the method. This will allow long-term analyses, which is important to examine the evolution of drusen in response to specific treatments. Furthermore, this computer aided evaluation will be faster than the manual one, which will give additional time to the clinicians in order to review the diagnosis and to follow-up the treatment.

The methodology that is proposed for the drusen automatic detection defines all image processing to determine the area affected by drusen, in order to establish a uniform analysis criterion, although it does not perform any automatic diagnosis. The retinal images are the only underlined data and therefore, no definitive diagnosis can be made, since complementary exams to other health indicators are usually required. Any result produced by an automatic lesion detection system is an exam that requires interpretation by a clinician, other exams and the patients' biological profile to be validated. As automatic diagnostic systems do not have this overall view of the patient, they can always produce false negative and false positive results.

The retinal images used in this work are complex. These contain anatomical structures that are in its surface, such as the optic disk, the vessels and lesions when they exist (drusen, exudates, haemorrhages and others), alongside with the high variability among different patients' images due to the differences in each individual's physiology. However, this

methodology concentrates only on the drusen and in particular on those that affect the patients' central vision. The other structures, vessels and optic disk, are also used in this methodology but as auxiliary indicators for contrast adjustment and to obtain the image resolution.

The automatic detection and quantification of drusen is described as a methodology composed by five steps, starting with the image pre-processing and finishing with the quantification of drusen. In this chapter, after presenting this methodology, the methods that were used for its' validation alongside the clinicians will be described.

3.1 - The Methodology

Although there are no strict rules on how a digital image processing system should be structured, it is usually divided into a sequence of four operations (Gonzalez and Woods 2007). It starts with a digital format *image capturing*, followed by three bottom-up sequences of operations for *image pre-processing*, *image interpretation* and *cognitive processing*. The *image pre-processing* is characterized by low-level image operations where the inputs and outputs are images, such as image enhancement, noise reduction, geometrical operations and others, in order to prepare and normalize the image for further processing. The *image interpretation* is a mid-level operation which describes the image contents. This involves image segmentation, to separate the image in regions or objects, object characterization and object recognition to identify the structures in the image. At this level the inputs are images and outputs are attributes extracted from the structures in the image. The *cognitive processing* on its turn is to use the information retrieved and process it accordingly. It can include analysis of sequences of images, alarms generation and other applications.

The methodology proposed for drusen automatic detection defines all image processing to determine the area affected by drusen in fundus images. The choice to analyze the fundus images was that they are the most widely available and used method to diagnose drusen. The *image capture* and the *cognitive processing* are out of the scope of this methodology. The images are already supplied in digital format ready to be processed, while the *cognitive processing* is the medical diagnosis itself.

Following the ophthalmologists recommendations involved in this work and the Wisconsin grading system (Klein, Davis et al. 1991), the analysis is focused only on the central vision, namely the inner macula, i.e., a circular region of interest (ROI) of 3000µm of diameter centred on the fovea. Within this region, drusen are identified and quantified, as

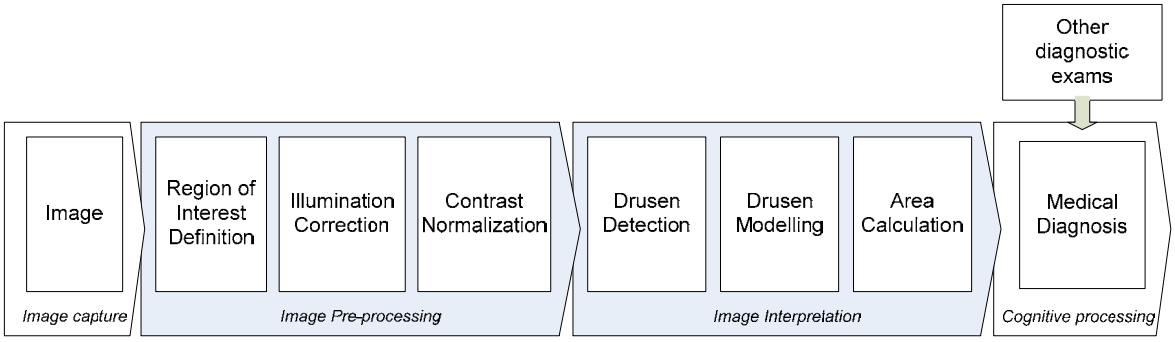


Figure 3.1 – Methodology for automatic drusen detection and quantification.

presented in the figure 3.1. The proposed methodology covers only the *image pre-processing* and *interpretation*.

Since the image acquisition conditions differ between Clinical Centres and equipments, the *image pre-processing* normalizes the image for the drusen detection and quantification process. This normalization will include both image intensity and contrast. Due to eye lens problems and to deficient patient collaboration during the image acquisition the illumination over the retina is in many cases non uniform. The *image pre-processing* is also responsible for correcting this non uniform illumination.

The *image interpretation* is composed by the detection of drusen locations followed by an analytical characterization and finally by the affected areas calculation. These sub-operations include: *Drusen detection*, *Drusen modelling* and *Area calculation*. For drusen detection a new algorithm for local maximums detection (Gradient Path Labelling) with improved accuracy and performance was developed. It retrieves the drusen location and amplitude, which are required as initial parameters to the modelling step. The modelling of drusen is obtained with a fitting of modified Gaussian functions for the detected drusen, allowing its analytical characterization. With the drusen model, the affected area, as well as other indicators such as the number of drusen or the overall reflectivity, can be finally quantified.

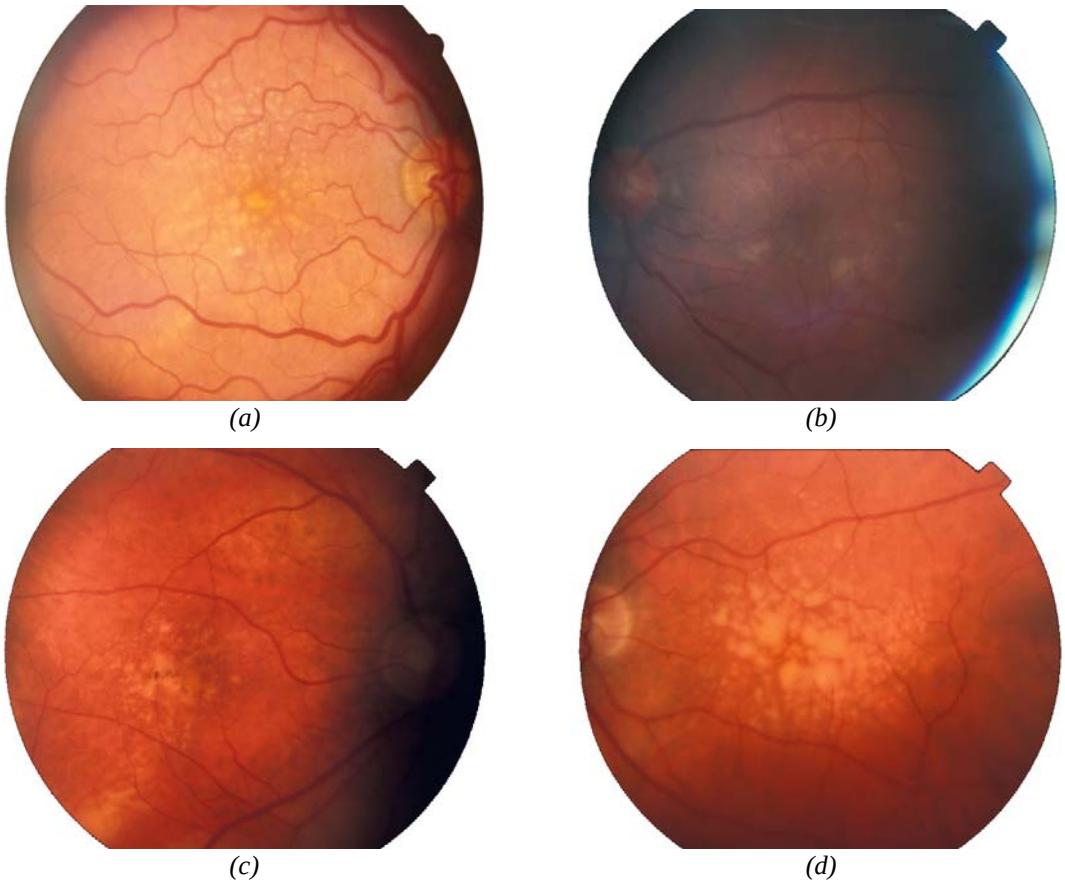
The parameterization in the automatic image processing systems is one of their weak points. If on one side a parameterizable system is capable of being adjusted to more types of images, on the other side it is less adaptive by requiring pre-saved configurations for each image type. Also, for the definition of a Gold-Standard drusen assessment a flexible method is needed, otherwise its status can be questionable. Having a Gold-Standard definition as a possible objective to achieve, in this work there was a concern to design a methodology with as less parameters as possible.

The procedure for setting the base parameterization is in a first phase to use images graded by the clinicians to provide feedback for the initial parameters. Then the algorithms are adjusted to produce similar results as do the specialists. As the consensus among the specialists is some times weak, as well as a moderate intra-operator variability, the comparison between the automatic detection algorithm and all the specialists should have a statistical approach to allow some degree of uncertainty in the model.

A second phase pursuing the development of the Gold-Standard is to present the automatic detection of drusen to the specialists in order to assess its results and adjust the parameters accordingly to achieve a consensus among them. With the final parameterization defined, the last phase is to validate the results with a blind study. Images are randomly presented both to the ophthalmologists and to the software in order to assess the agreement.

3.2 - Image Pre-processing

The main objective for the image pre-processing is to improve the contrast between drusen and background and to normalize the image conditions which the posterior methods will use. The image's illumination, colour and alignment changes, between the retinograph



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

Figure 3.2 – Examples of retinal images acquired with different conditions. (a) Normal; (b) Deficient illumination; (c) (d) Non-uniform illumination.

equipments and supports used (digital or analogue), require an efficient pre-processing to achieve normalization. figure 3.2 illustrates some examples of retinal images that result of different acquisition conditions, where one normal image and three images containing deficient illumination and non-uniform illumination are presented. The type of equipment and illumination used also introduce differences in the captured image which should be preferably solved by the *image pre-processing* phase.

The figure 3.3.(a,b) shows retinal examples of a analogue contrast improvement method which consists in illuminating the retina with red-free light. The produced image can be in a grey-scale or in a greenish scale. By using a non-mydriatic camera, the retinal images will have a slight different coloration as it is presented in figure 3.3.(c,d), although the changes in contrast are not significant.

Although the detection algorithms are not influenced by the image orientation, for results comparison the fundus images should be orientated, as they already are, before applying the algorithms. Also to compare sequences of images, these should be properly registered before the drusen detection.

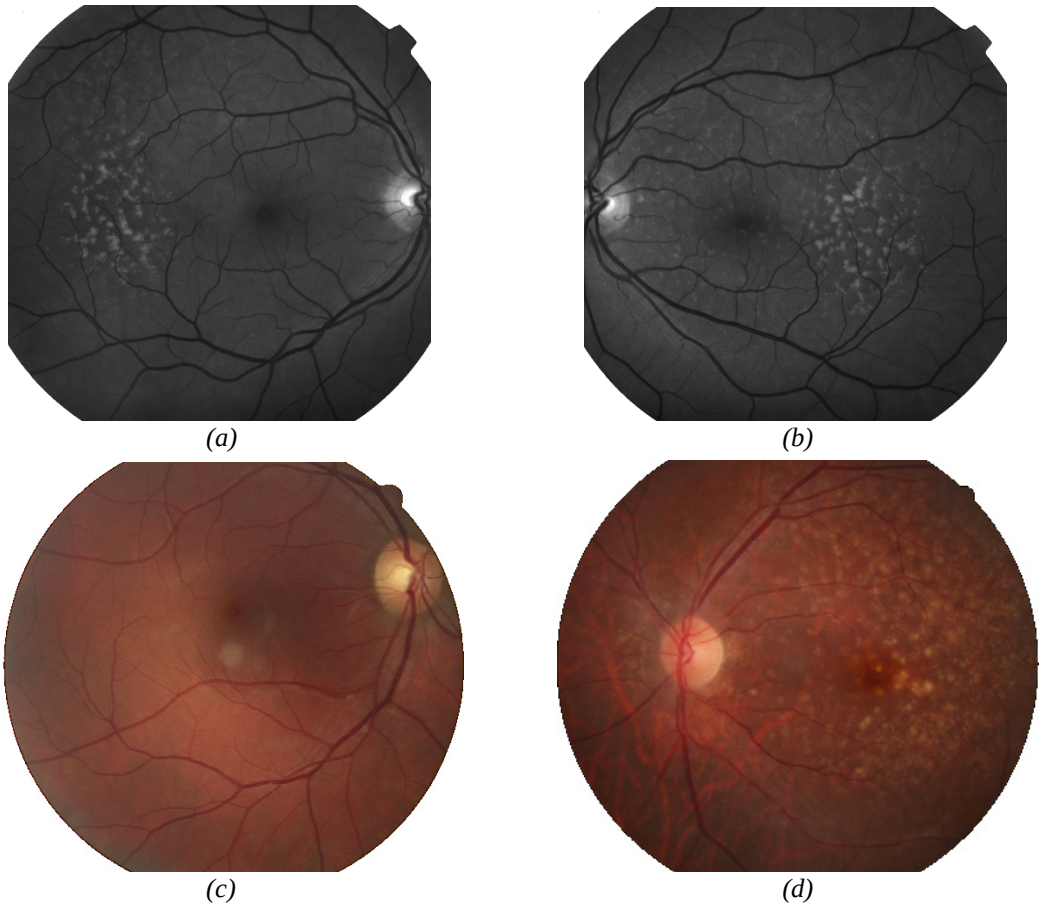


Figure 3.3 – Retinal images with analogue contrast Improvement. (a) (b) Retinal images acquired with red-free illumination. (c) (d) Images acquired from a non-mydriatic retinograph.

The image resolution is also important to be normalized. It will influence for example the use of digital filtering by requiring variable size kernels, the choice for the detection thresholds and also the systems performance. However, in this work it was not needed as the algorithms used were prepared to cope with this variability by adjusting the parameters accordingly to the image's resolution.

The *image pre-processing* proposed in this thesis will then be composed by the region of interest definition, the colour selection, the correction of the illumination non-uniformity and the contrast normalization.

3.2.1. Region of interest definition

The first operation over the image is to locate the region of interest (ROI) where the image operations will produce effects and where drusen are to be detected and quantified. This is done by locating the macula and specifying a circular ROI around it. The fundus images are usually already oriented and aligned. However, to verify the alignment, most of the retinographs add an orientation mark on their top-right side (figure 3.4.a). With this orientation it is possible for the clinicians to orientate the image and identify which eye it is. A right eye has its Optic disk on the same side as the orientation mark, while the left eye has the opposite.

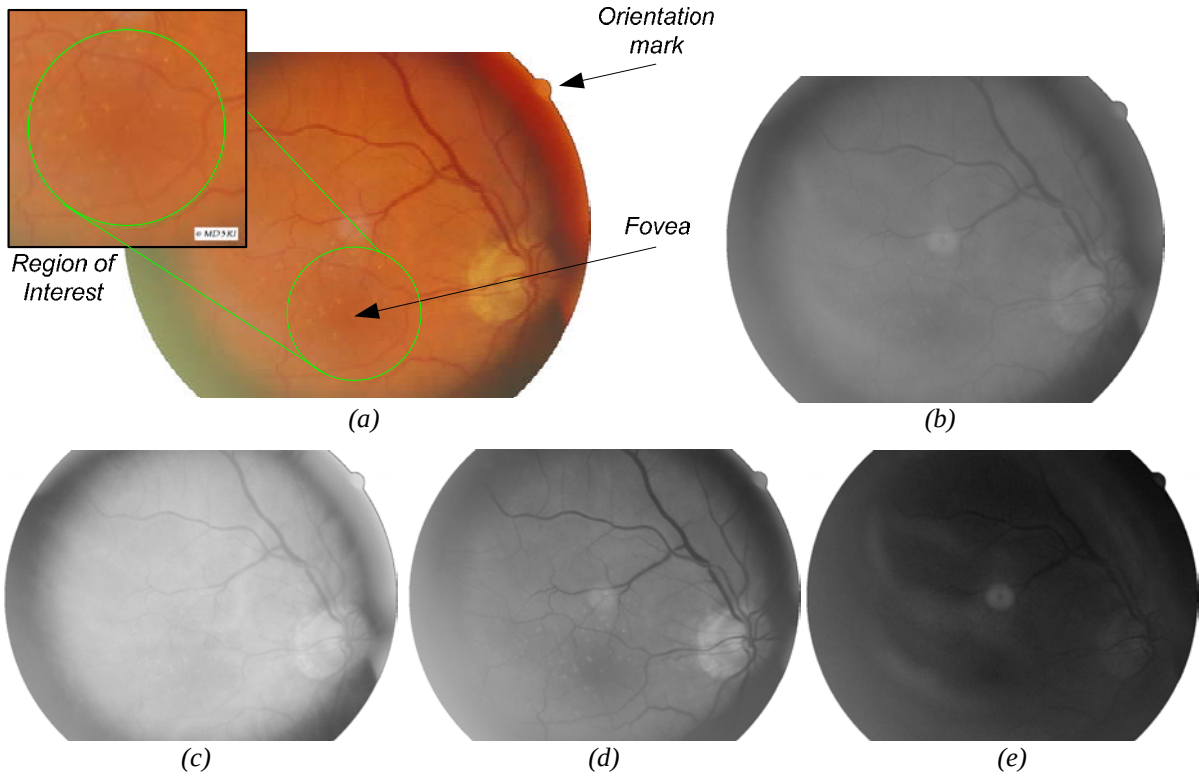


Figure 3.4 – Retina image description. (a) RGB image with annotations and region of interest zoom; (b) grey channel; (c) red channel; (d) green channel; (e) blue channel.

In the images presented it is noticed a darker area near the centre in 3 to 4 mm to the temporal side of the optic disk, which is the macula. It can also be identified by the convergence of the retinal vessels in the centre (figure 3.4.a). This latter is an important reference in images containing advanced macular degeneration, since the macula may gain some clarity and make difficult its location.

The ROI will be a circular area centred on the macula with a diameter of two optic disk diameters (ODD). The ODD is used in several studies as a standard reference to calculate the image resolution. Despite that it is a stable measure among the humans (Jonas, Gusek et al. 1988), there is no consensus within the scientific community in the reference value, ranging from 1500µm to 1850µm. In this work a reference value of 1500µm is used according to the International grading recommendations (Bird, Bressler et al. 1995).

Although there are some works relating to the automatic macula and optic disk location (Hoover and Goldbaum 2003; Tobin, Chaum et al. 2007), they are not in the scope of this work. These works use the vessels and optic disc geometric configuration in order to estimate the exact location of the macula (figure 3.5). The work from Tobin *et al.* (2007) achieved an accuracy higher than 90%, in a set of images containing several retinal pathologies. Despite these good results, the work being presented is focused in drusen detection, leaving for the specialist the manual detection of the macula, which executes a similar cognitive procedure as the one showed in the Figure. First, they measure the ODD in pixels and then mark a circle with two ODD centred on the macula.

After defining the ROI, the image is cropped to a square centred on the ROI with a width of 1.2 x 2 x ODD, in which the factor 1.2 is the margin used to avoid a special image

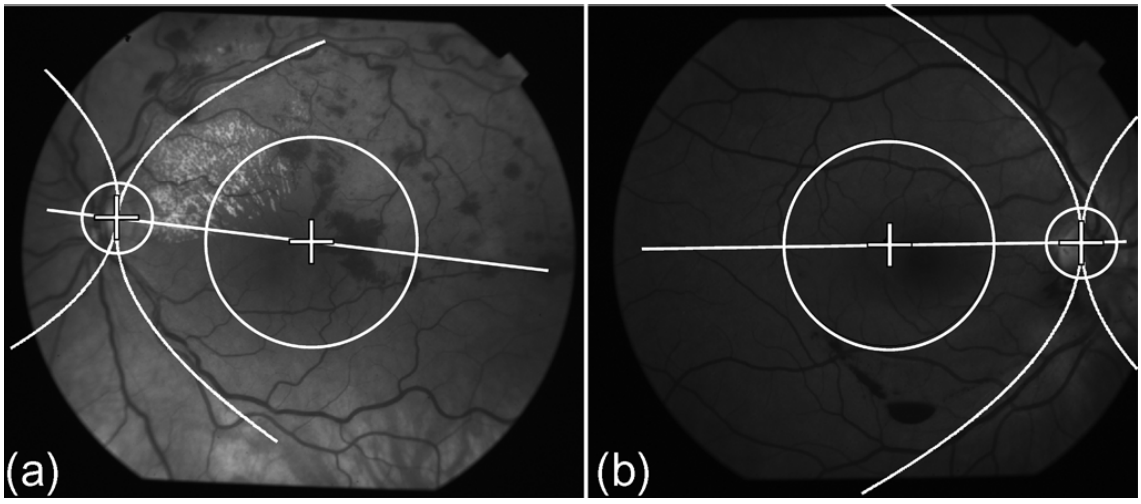


Figure 3.5 – Locating of the macula using the geometric configuration of blood vessels and optic disc. (*from* (Tobin, Chaum et al. 2007))

processing in the ROI boundaries. The resulting ROI, which is zoomed in figure 3.4.a, is then used in the following image processing step.

3.2.2. Colour channel selection

From the analysis of the image colour channels, as presented in section 2.2.1, and despite the fact that the images are predominantly red, it was noticed that for colour images the green channel is the one providing the best contrast and which is less affected by the illumination non-uniformity. In figure 3.4.(b,c,d,e), which shows the individual colour and grey channels, it is again noticed that the blue channel has almost no information, the red channel is the one with higher intensity and with eventual saturation, the grey channel has a low and non-uniform contrast, and the green channel has a medium intensity and good contrast.

The image histograms presented in figure 3.6 confirms these observations. The red channel has expanded the intensity axis as a result of the non-uniform illumination and has a higher concentration on brighter intensities, being possible to mistake with drusen intensity. The grey channel exhibits a good histogram shape, although the image contrast is not uniform. From the four histograms presented, the green channel is the more balanced. It has a peak value near the centre, representing the background, and has clearly defined secondary peaks corresponding to other structures as the vessels, the optic disc, the drusen and others.

The use of the hue channel from the IHS colour space is also an alternative which has also been referred in section 2.2.1. However, the inconsistency of the image colour between the imaging techniques makes it a less reliable approach. Therefore, due to the consistency and to the less influence of non-uniformity of illumination, the green channel was the chosen one to be used in further processing without any additional transformation.

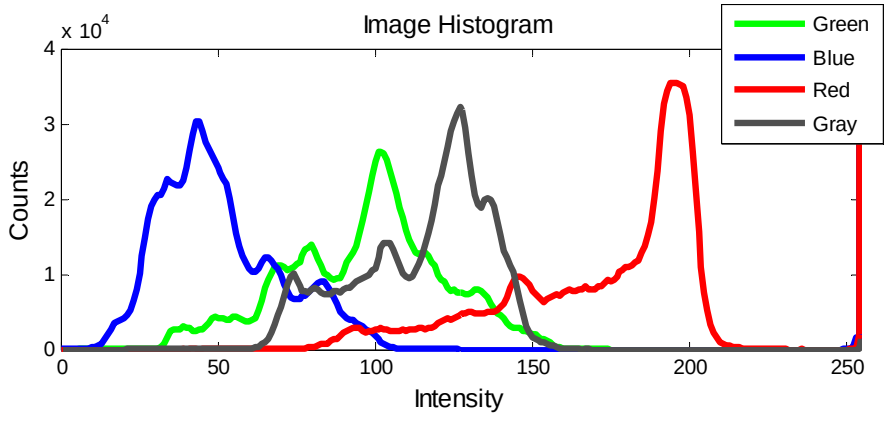
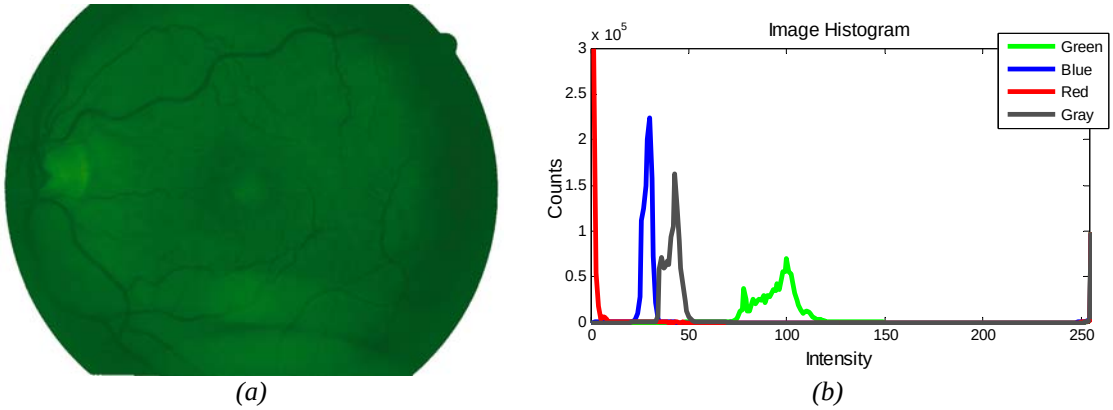


Figure 3.6 – RGB and Grey histograms of the retina image in figure 3.4.a.



(a) *(b)*

Figure 3.7 – Red-free light illumination. (a) Retina image; (b) colour histogram.

The retina image illuminated with red-free light as shown in figure 3.7.a, uses the same procedure. Here the red channel has some residual information; the blue channel is also non-informative, being only the green channel the one with a good contrast and a balanced histogram (figure 3.7.b).

3.2.3. Non-Uniform Illumination Compensation

In ophthalmologic imaging it is usually difficult to obtain good quality images. *Illumination* is not always uniform over the retina due to numerous factors, such as, the retina convex surface, the lower macula reflectance in comparison to others parts of the retina, the existence of ocular pathologies which affect the lens, poor patient collaboration, among other factors. The compensation of non-uniformity of illumination is perhaps one of the key operations in this *image pre-processing*. This illumination effect is responsible for the appearance of brighter and darker areas and different contrasts among them. The purpose of this pre-processing operation is to normalize the illumination, in order to have uniform intensity and contrast patterns along the retina and especially in the macula.

The illumination pattern that focuses on the retina is not known *a priori* since it is dependent on various unknown independent interferences. Therefore, it should be estimated based on the acquired image. Two approaches for correcting illumination were analysed, one based on frequency filtering and other based on illumination estimation and compensation, which will be presented next.

3.2.3.1 Frequency Filtering

The non-uniform illumination pattern that is visible in fundus imaging has typically smooth light variations. When analysing it in the frequency domain it is characterized by the presence of low frequency components. These observations conduced to the evaluation of the

effectiveness of a high-pass filtering operation in order to reduce the illumination non-uniformity.

Frequency domain filtering

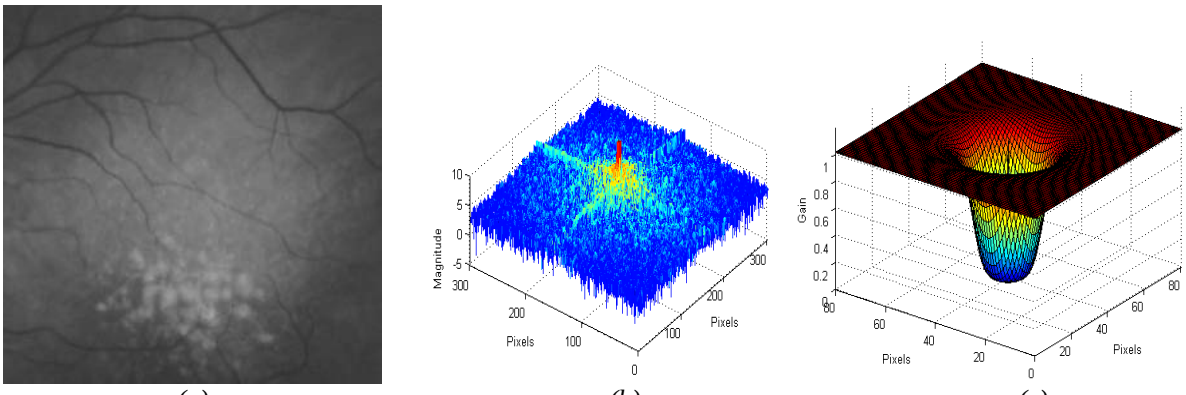
Image processing in the frequency domain is the basis of linear filtering with more flexibility in the design and implementation of filtering solutions. By applying a 2-D Fast Fourier Transform (FFT) to an image the magnitude and the angle planes of the frequencies are obtained. In the frequency domain, by weighing the frequencies magnitude, filtering operations are implemented. A low-pass filter will reduce the magnitude of the higher frequencies, while the high-pass performs the opposite. Unless the filter design wishes to change the image orientation the angle plane is not used for filtering.

The concept supporting linear filtering in spatial domain is the convolution theorem. Here the filter mask, *h(x,y)*, and the image, *f(x,y)*, are convoluted and generate a filtered image. Through the convolution theorem, in the frequency domain this operation can be written as:

$$f(x,y) * h(x,y) \Leftrightarrow F(u,v) \cdot H(u,v) \tag{3.1}$$

where * denotes convolution, *F(u,v)* the image FFT and *H(u,v)* the filter transfer function in the frequency domain. Therefore, to implement a filter in the frequency domain, the filter transfer function must be designed to perform the necessary frequency modifications, and multiply it by the image FFT. Applying the FFT inverse transformation, the filtered image is obtained in the spatial domain.

In figure 3.8.(a,b) the FFT magnitude for an image with non-uniform illumination is presented. It is noticed a higher magnitude of the lower frequencies at the centre which are responsible for the illumination. To compensate the non-uniformity it was designed a high-



(a) *(b)* *(c)*

Figure 3.8 – Image processing in the Frequency domain. (a) Original image; (b) FFT magnitude plane; (c) high-pass filter transfer function (zoomed to 0..80 pixels).

pass Butterworth filter which removes the contribution from the lower frequencies (figure 3.8.c). This filter has two configurable parameters to define the cut-off frequency and the filter order shape.

To evaluate this filtering approach two examples are presented in figure 3.9, which compare two different filter configurations. Lines (a,c) contain the configuration which produced better results; the illumination is uniform and the contrast between background and other structures are maintained (vessels and drusen). Lines (b,d) have a higher cut-off frequency which has over-filtered the image, the illumination is uniform, but the contrast was reduced. The images (c,d) contains a large density of drusen, and it is visible that, although the image background is uniform, the contrast is not. It has lost some of the contrast in the areas with more drusen and in the previous darker areas the contrast is poor.

This filtering technique is promising to correct the background, although it does not correct the non-uniformity of contrast. Also, the configuration of cut-off frequency is dependent on the image resolution and on the severity of the pathology.

3.2.3.2 Homomorphic filtering

The previous technique showed that it is not enough to normalize the image background to have uniform intensity and contrast, once it has normalized the intensity but not the contrast. To overcome this problem, the *homomorphic filtering* technique was tested. As it was mentioned on section 2.2.2 this technique is used to separate illumination and reflectance in the spatial domain and correct the non-uniformity of illumination in frequency domain (Gonzalez and Woods 2007).

The retina is said to be a *Lambertian* surface (Narasimha-Iyer, Can et al. 2006); a diffuse reflecting surface where the light reflected from the surface is directly proportional to the *cosine* of the angle θ between the observer's line of sight and the surface *normal*. Assuming that illumination is approximately a perpendicular to the retina surface ($\theta=0^{o}$), the pixel values, and consequently the contrast, vary accordingly with the amount of illumination $i(x,y)$ and surface reflectance $r(x,y)$.

These two functions are combined as a product to form $f(x,y)$:

$$\begin{aligned} f(x,y) &= r(x,y)\cdot i(x,y)\cdot\cos(\theta) \\ \Rightarrow \text{assuming } &\theta \approx 0^{\circ} \\ f(x,y) &\approx r(x,y)\cdot i(x,y) \end{aligned} \tag{3.2}$$

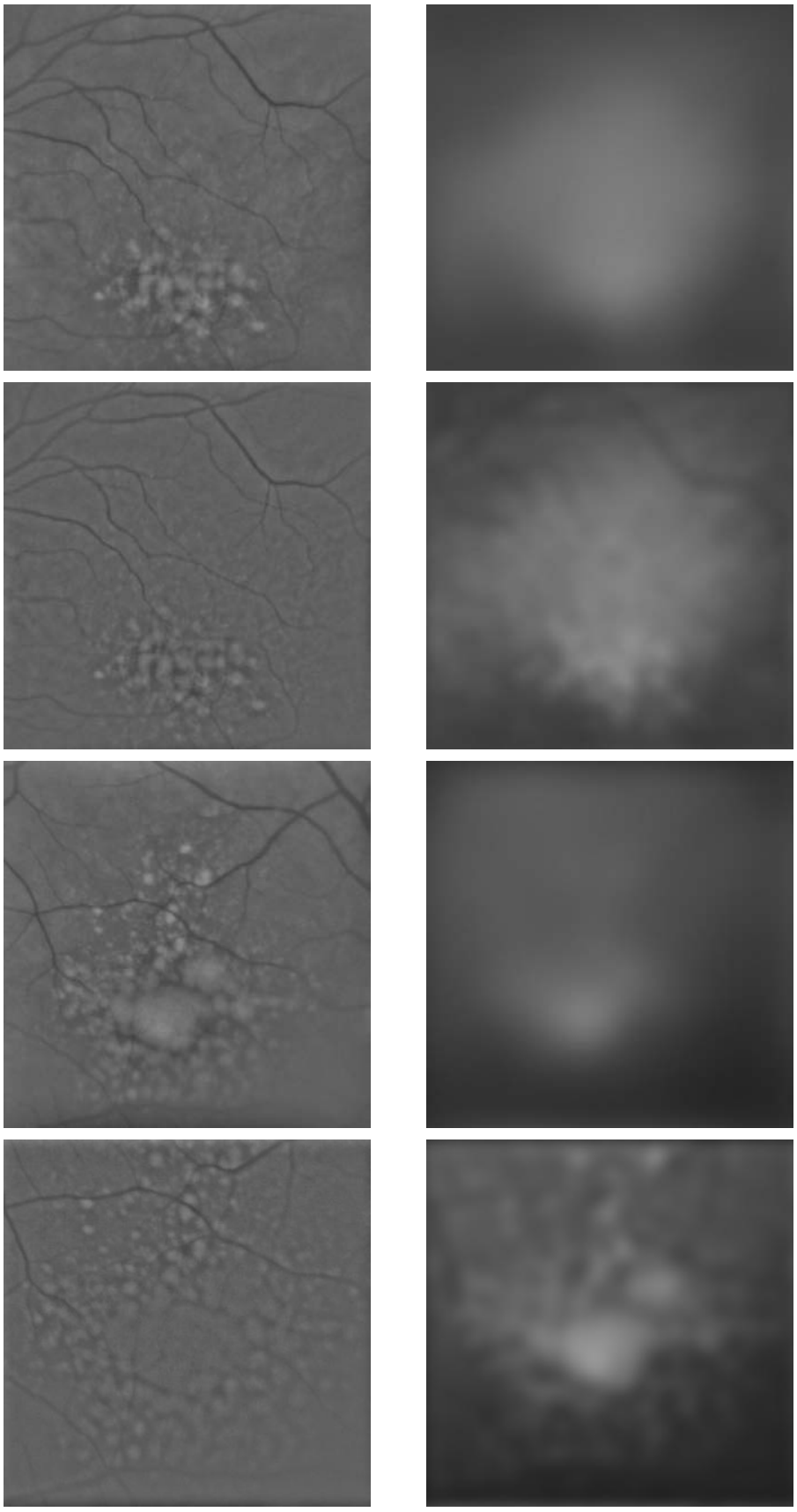


Figure 3.9 – Filtering in the frequency domain using a Butterworth high-pass filter with two configurations: (a,c) filter order = 2 and cut-off frequency = 4 and (b,d) filter order = 3 and cut-off frequency = 12. On the left are the results from applying the *high-pass* filter and on the right are the results from applying the *low-pass* filter.

Based on the fact that the image contrast is proportional to the illumination, the *homomorphic* technique separates the illumination from reflectance, and by correcting the illumination non-uniformities, it is obtained a normalized intensity and contrast. The procedure starts by converting the image into the logarithmic domain where the multiplicative factors are made additive.

$$\ln\big(f(x,y)\big) = \ln\big(i(x,y)\big) + \ln\big(r(x,y)\big) \tag{3.3}$$

Using the same technique presented before, the low frequencies observed in the illumination pattern are removed in the frequency domain. The linearity properties of the inverse Fourier Transform will transform the illumination removal on a negative logarithmic member, which in the linear domain will generate a division term (3.4). This operation, by generating a normalized illumination, will produce a normalized contrast also.

$$\ln\big(f_{corrected}(x,y)\big) = \ln\big(i(x,y)\big) + \ln\big(r(x,y)\big) - \ln(i_{removed}(x,y))$$

$$f_{corrected}(x,y) = \frac{i(x,y)\cdot r(x,y)}{i_{removed}(x,y)} = i_{normalized}(x,y)\cdot r(x,y) \tag{3.4}$$

In figure 3.10 are shown the same examples as figure 3.9 but after applying the homomorphing technique with a Butterworth filter (*cut-off* = 2, *order* = 2). It can be observed that illumination is uniform as it is in normal filtering and that the contrast is uniform and improved. This is especially noticed on the bottom region of figure 3.10.b where the drusen stand-out.

This is an effective technique for normalizing the image background and contrast. By using the basis of the illumination theory, separating the illumination from reflectance and correcting the first, it not only corrects image background but also normalizes the contrast. The configuration of the filter's cut-off frequency and amplification factor to obtain normalized images are the main drawbacks of this technique.

3.2.3.3 Illumination estimation and compensation

Based on the experience obtained with the previous technique, other techniques that estimate the illumination pattern and replace it by a uniform pattern were exploited. They make use of illumination theory and of the equation (3.2) to adjust illumination to a constant pattern. What is proposed is to estimate the non-uniform *illumination* function $i_{estimated}(x,y)$ and through the equation (3.2) obtain the surface *reflectance r(x,y).* As a consequence, only

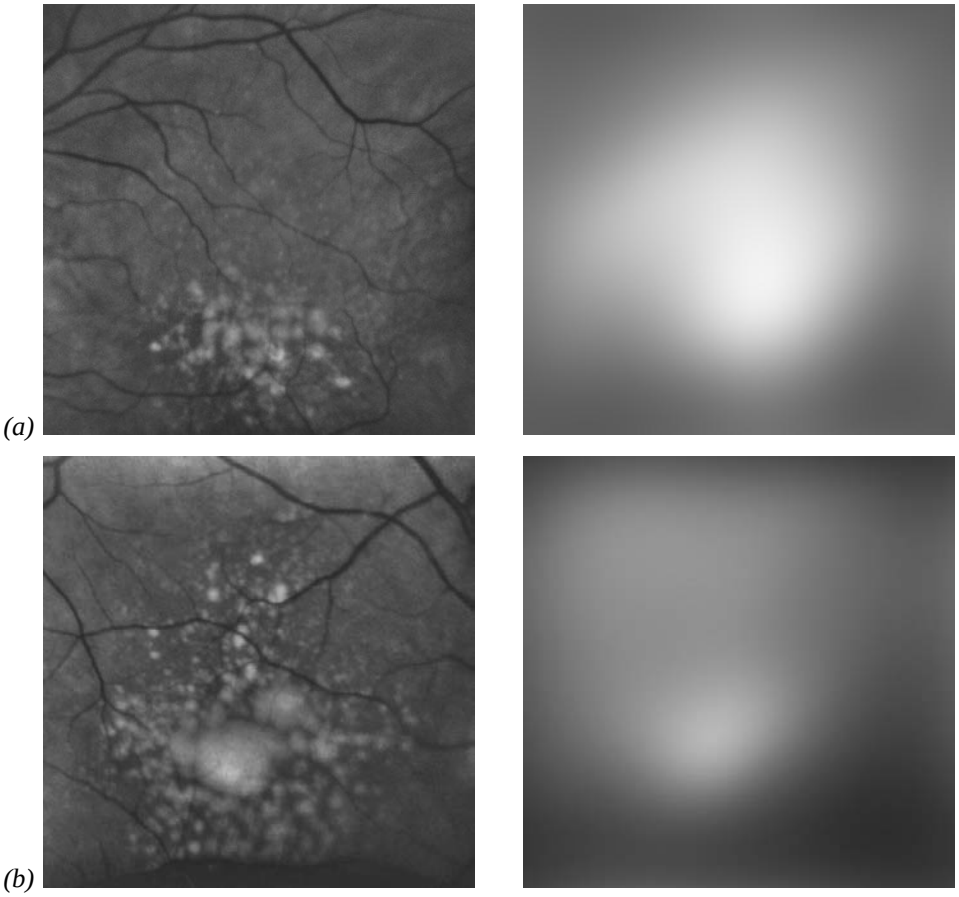


Figure 3.10 – Examples of the *homomorphic* filtering technique. (a) image containing medium sized drusen; (b) image containing large drusen.

the retina surface characteristics, without non-uniform *illumination*, are obtained, being multiplied by a constant illumination, from which results the normalized image.

$$f_{corrected}(x,y) = \frac{f(x,y)}{i_{estimated}(x,y)} \cdot constant\_illumination \qquad (3.5)$$

Three techniques for estimating the illumination pattern were studied. They differ in the algorithms' processing weight and in the capacity to estimate complex illumination patterns.

Gaussian fitting

The first exploited algorithm was based on the modelling of illumination using Gaussian functions. The smooth characteristic of the Gaussian function showed to be well adapted to represent light distributions (Driggers 2003). Therefore, to model the illumination pattern three high width Gaussian functions were empirically chosen, representing three superimposed light sources coming from different locations with different sizes and shapes. Other illumination models with more or less functions are possible. However, they will

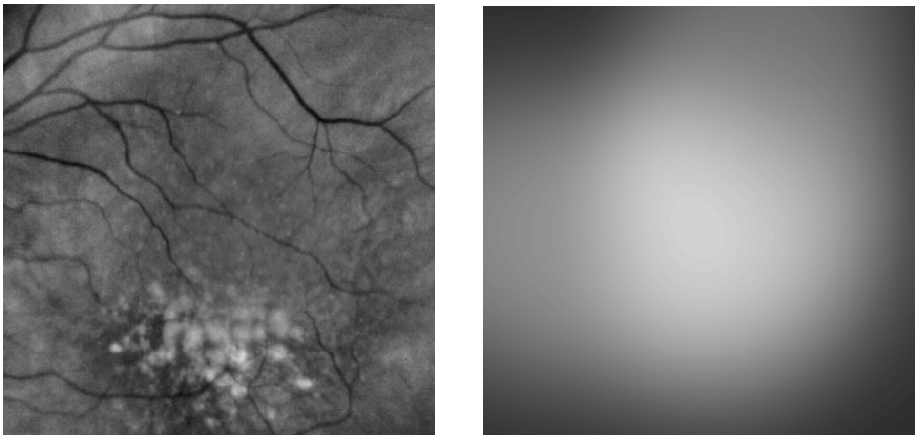


Figure 3.11 – Illumination compensation using Gaussian fitting.

overestimate or underestimate the illumination pattern, which is supposed to be smooth and have small variations.

The Gaussian functions were modelled to the image by the Levenberg-Marquardt fitting algorithm (Marquardt 1963; William, William et al. 2002). This algorithm adjusted the model function parameters in order to minimize the chi-square value between the original image and the estimated illumination model. Then, by dividing the generated model from the original image, the normalized image was obtained.

In figure 3.11 is presented a single example of the application of the technique. It provides a good estimation of the illumination pattern, although it is difficult to parameterize in order to guarantee the convergence of the fitting algorithm. The model's specification, i.e., the number of functions, their initial parameter settings and their limits are also factors that forced the rejection of this method for correcting illumination.

Gaussian blurring

The Gaussian Blurring algorithm (Sedgewick 2008) is a technique which extracts the illumination pattern through the application of a low-pass filter in the spatial domain followed by the illumination normalization using equation (3.5). In order to estimate the illumination it is used a large sized Gaussian filter, having about 1/3 of the image size, which extracts the low frequency changes.

The filter uses a mask calculated by the equation (Gonzalez and Woods 2007):

$$GB(x,y) = \frac{100}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.6}$$

In this equation the parameter r represents the mask dimension (r x r mask) and the parameter σ is the Gaussian standard deviation. This latter was indexed to 1/3 of the mask

width to be wide in order to cover the low frequency changes. The illumination pattern is obtained by convoluting the filter mask with the original image. Finally, the normalized image is the result of applying this pattern in the equation (3.5).

In figure 3.12 are presented two examples of illumination compensation using the Gaussian Blurring algorithm. As it can be observed, the technique is well suited for images containing medium sized drusen. In the retinas with large sized drusen (*b*) the centre of the confluent drusen had a slight reduction of contrast, but the overall result is a normalized illumination and contrast. The filter mask can be adjusted to have a lower cut-off frequency and improve the behaviour in these cases; however this is a compromise in any situation. One disadvantage is the computation required for the filter convolution. For example, an image identical to the one shown has a filter size of 150x150 pixels wide. Therefore, for every pixel in the image (typically 450x450) it is needed to compute the convolution, which is time consuming. Some strategies as the image down-sampling or the use of one-dimensional kernel that calculates the horizontal, vertical and other two or more directions independently, reduce computation time.

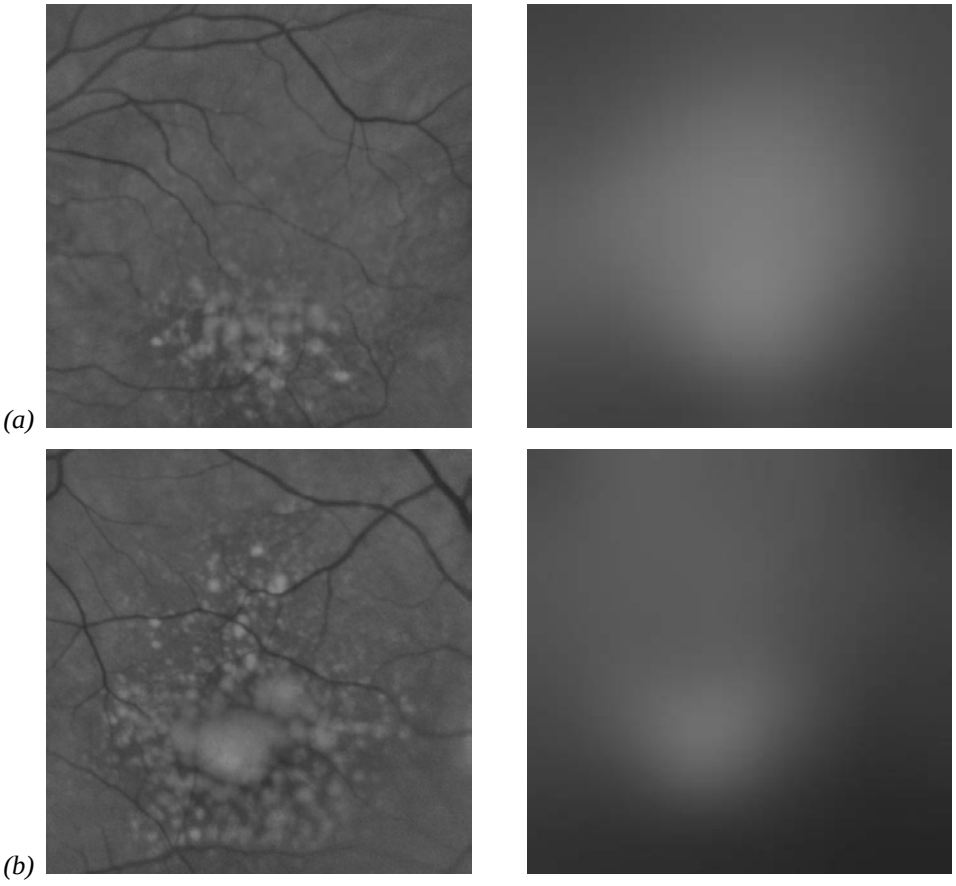


Figure 3.12 – Examples of application of the Gaussian Blurring Algorithm. (a) image containing medium sized drusen; (b) image containing large drusen.

Smoothing Spline fitting

Other technique for illumination estimation is to make use of the interpolation and smoothing characteristics of two-dimensional *Splines* curves to describe the illumination pattern. *Splines* are piecewise polynomials that are commonly used for interpolating data in one-dimension or multi-dimension. They are a sequence of polynomial curves without discontinuities in their edges. The most common are the Cubic *Splines* with continuity in the second derivative; this forces the *Spline* to be a straight line on the borders of the interval, while not disrupting its smoothness.

To estimate the illumination pattern the image background *f(x,y)* was fitted using the Cubic Smoothing *Spline*. These are a special class of *Spline* which have a roughness factor to control the ratio between a straight fitting of the data and a smooth *Spline*. The fitting algorithm for the *Spline*, *S(x,y)*, is to minimize the function presented in equation (3.7), where *p* is a smoothing factor and $D^2S(x,y)$ is its second derivative:

$$p \cdot \sum_{y=1}^{m}\sum_{x=1}^{n}\left|f(x,y)-S(x,y)\right|^2 + (1-p)\iint\left|D^2S(x,y)\right|^2 dxdy \qquad (3.7)$$

The smoothing factor *p*, defined in the interval [0..1], controls the relative weight between fitting all data points (*p*=1) and a smooth spline (*p=0*) (see figure 3.13). For *p=1* the spline is an interpolating spline that crosses every data point, while for *p=0* it minimizes the integral of the second derivative, i.e., maximizes the smoothness generating a straight line. The image resolution also influences the smoothness. As the resolution increases the pixels' spacing decreases and the *p*'s value should be increased to maintain the same smoothness.

To choose an adequate smoothing factor, experimental tests were carried out on a set of 22 images. For an image resolution of 12.5 *µm/pixel* a reference value of $p = 1e^{-6}$ was obtained. To find the correlation between the smoothing factor and the image resolution the

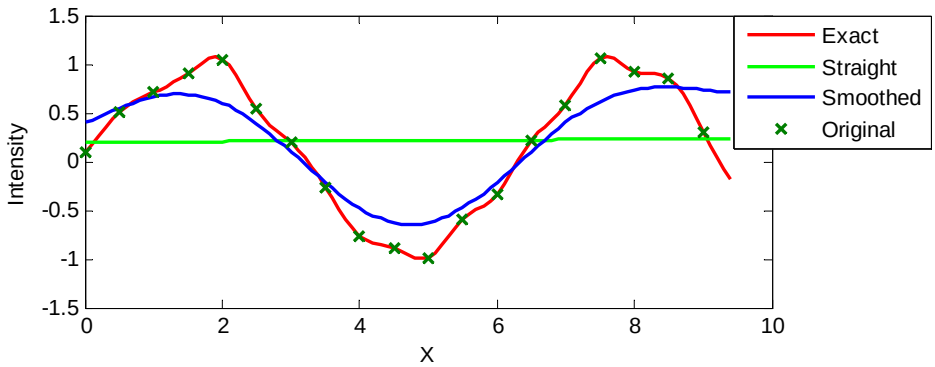


Figure 3.13 – Smoothing *Spline* Fitting examples for different *p* values.

test images were resized to predefined scale factors and new p values were estimated to produce the same smoothing effects. The comparison was made by resizing the reference *Splines* and by calculating the sum square error between these and the new smoothing *Splines*. The analysis of the smoothing factors showed a polynomial distribution as presented in equation below (consider x as the image resolution).

$$p = a_4 \cdot x^4 + a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0$$
$$\left\{ \begin{array}{ll} a_4 = 1.198e\text{-}010 & a_1 = -4.346e\text{-}007 \\ a_3 = -4.245e\text{-}009 & a_0 = 9.342e\text{-}007 \\ a_2 = 6.826e\text{-}008 & \end{array} \right\} \quad (3.8)$$

Figure 3.14 presents the results of the smoothing spline fitting on two examples used previously. The images on the left-side were obtained dividing the original image by the estimated smoothing spline. The results are comparable to the previous technique although its performance is faster and is less complex to derive. For the images containing large sized drusen the contrast on these confluent drusen is decreased. In these cases the large bright area provoked by drusen influences the *Spline* estimation being considered an illumination

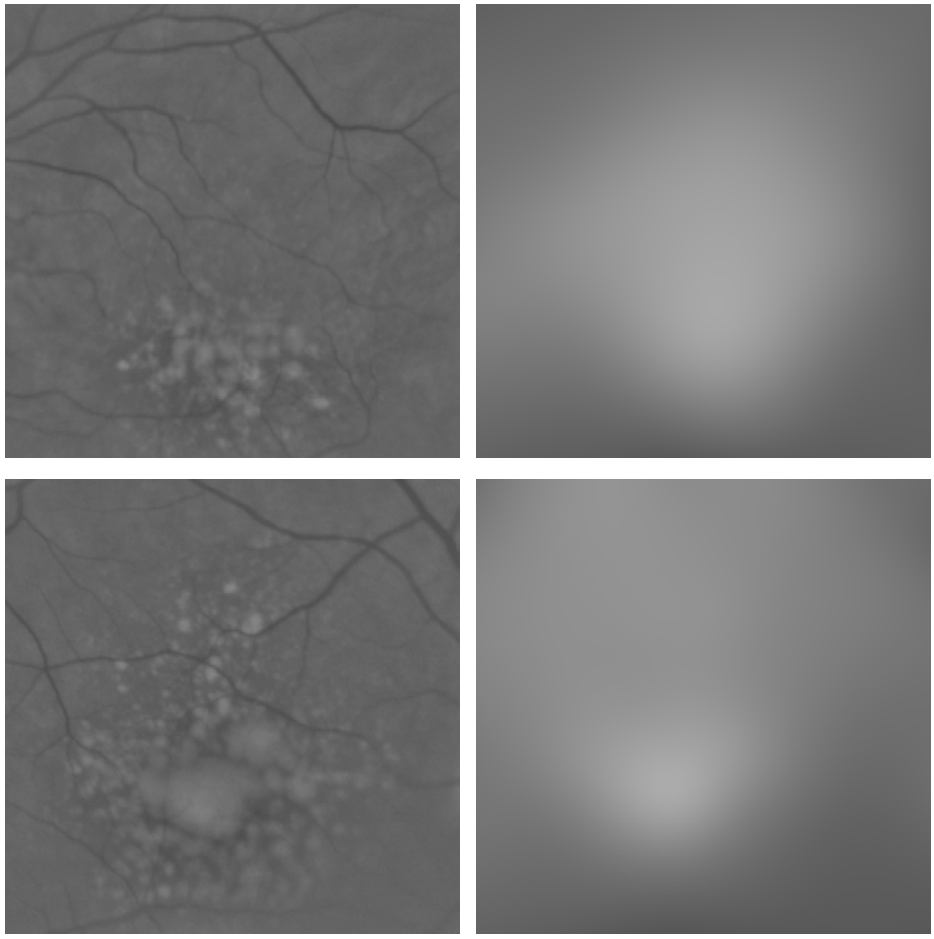


Figure 3.14 – Smoothing *Spline* Fitting Algorithm with a single pass.
(a) image containing medium sized drusen; (b) image containing large drusen.

deficiency. To improve the algorithm's efficiency an iterative method was designed to overcome the influence of drusen on the illumination estimation.

Iterative Smoothing Spline Fitting

To avoid drusen influence on the background estimation, an iterative process that removes drusen affected areas was implemented. The principle is to estimate the illumination pattern gradually after replacing the drusen bright areas by the previous estimated illumination. This will produce a more reliable illumination pattern less influenced by the large confluent drusen areas. However, there is a disadvantage to this process, on images containing severe non-uniform illumination the algorithm will fail when removing these areas from the illumination pattern. In these cases a user validation will be required. The iterative procedure (algorithm 3.1) is as follows:

1. Estimate and correct illumination using the Smoothing *Spline* Fitting algorithm (figure 3.15.(a,b)). This step is identical to the previous algorithm.
2. Cluster the pixels of the corrected image into two categories (drusen and background) using the Otsu thresholding method.
3. Replace the pixels belonging to the drusen class by the illumination estimated.
4. Repeat the three previous steps on the image that results from *step 3* until a predefined number of iterations take place or when the difference between consecutive iterations is below a predefined value.

Algorithm 3.1 - Iterative Smoothing Spline Fitting Algorithm

```
image_flatten = Function NonUniform_Illumination_Correction(original_image) {

    #define MaxIterations 5
    #define MinDifference 0.25

    working_image = original_image;
    oldbackSpline = original_image;
    iteration = 0;

    while (iteration < MaxIterations or splineDifference < MinDifference)
    {
    // step1
       backSpline = SmoothingSplineFitting(working_image);
       image_flatten = (original_image / backSpline) ;

    // step2
       image_cluster = OtsuThreshold (image_flatten);

    // step3
       working_image = ReplaceDrusen(image_flatten, image_cluster, backSpline);

    //step4
       splineDifference = (oldbackSpline – backSpline)/number pixels;
```

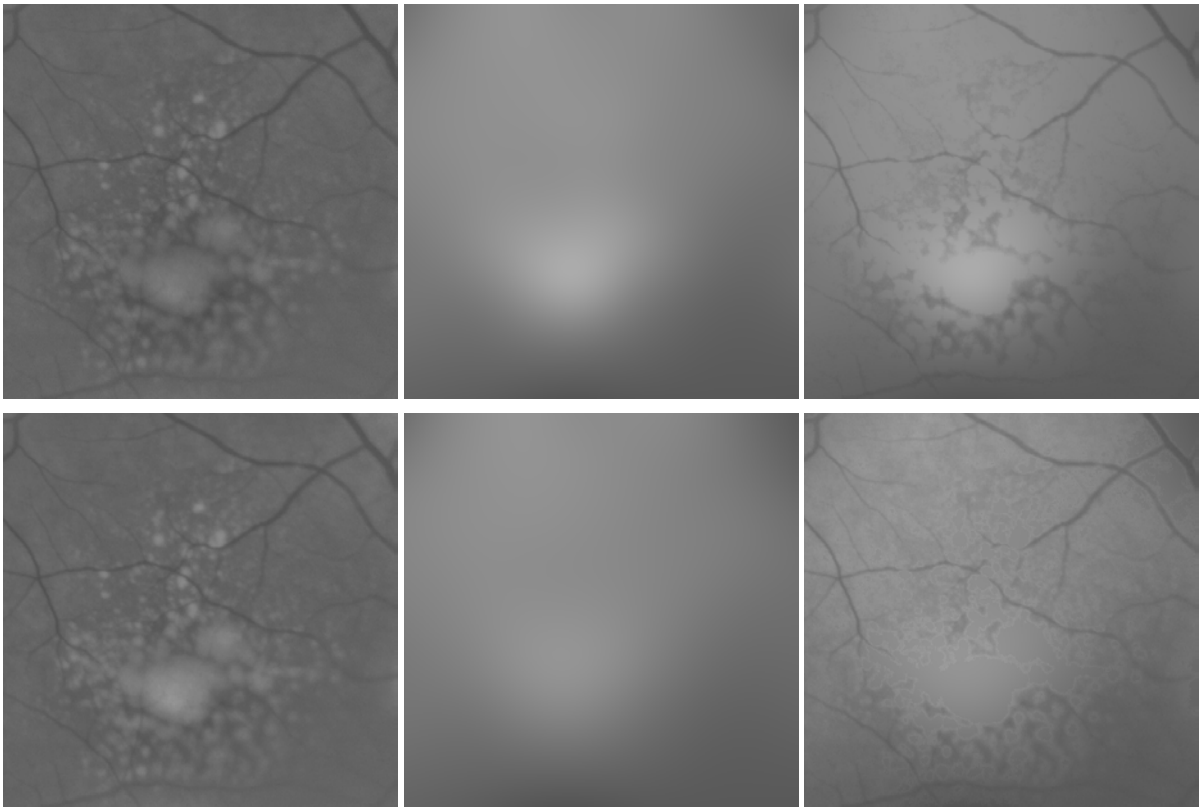
```
24:             oldbackSpline = backSpline;
25:             iteration = iteration +1;
26:         }
27:     }
```

As it is visible on figure 3.15 and figure 3.16 the contrast on drusen areas was improved, avoiding the influence of drusen on the illumination estimation. This procedure works also with flat images since the combination of the Otsu and the replacement operations generates always a normalized image.

3.2.3.4 Algorithms quantitative comparison

In order to choose the most adequate non-uniform illumination correction algorithm these images were evaluated under three different parameters: the uniformity of illumination, the contrast between the background and the drusen and the morphological similarity between drusen parts in the original image and its correspondent corrected image.

Several correction algorithms were applied to an image set containing 22 images which were evaluated under the previous three parameters and then averaged. It was assessed:



(a) *(b)* *(c)*

Figure 3.15 – Iterative Smoothing Spline Fitting over an image containing large confluent drusen. The first line contains the first iteration of the algorithm, while the second line contains the fifth iteration.
From left to right: the flatten image, the estimated *Spline* and the next working image.

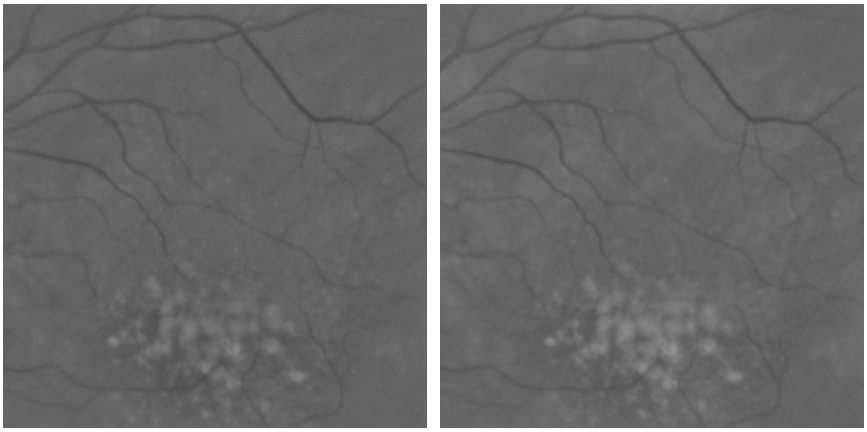


Figure 3.16 – Iterative Smoothing *Spline* Fitting over an image with medium sized drusen.

the original green image, FFT filtering with two filter configurations, Homomorphic filtering, Gaussian blurring, Smoothing Spline fitting and Iterative Smoothing Spline fitting. Before obtaining the parameters values, the images needed to be normalized in terms of intensity and contrast as well. A square region containing background and vessels was manually selected as the reference region where these last two parameters should be equivalent in all corrected images and in the original one. An addictive factor and a multiplicative factor were applied to correct illumination and contrast, respectively.

The uniformity of illumination parameter was assessed calculating the standard deviation of the background pixels' intensity. Since the detection of background pixels is a difficult task to automate accurately, the user manually selected several small square regions (30 pixels wide) where the mean intensity was computed (orange squares on figure 3.17). There were at least 8 regions and were dispersed along the whole image. The standard

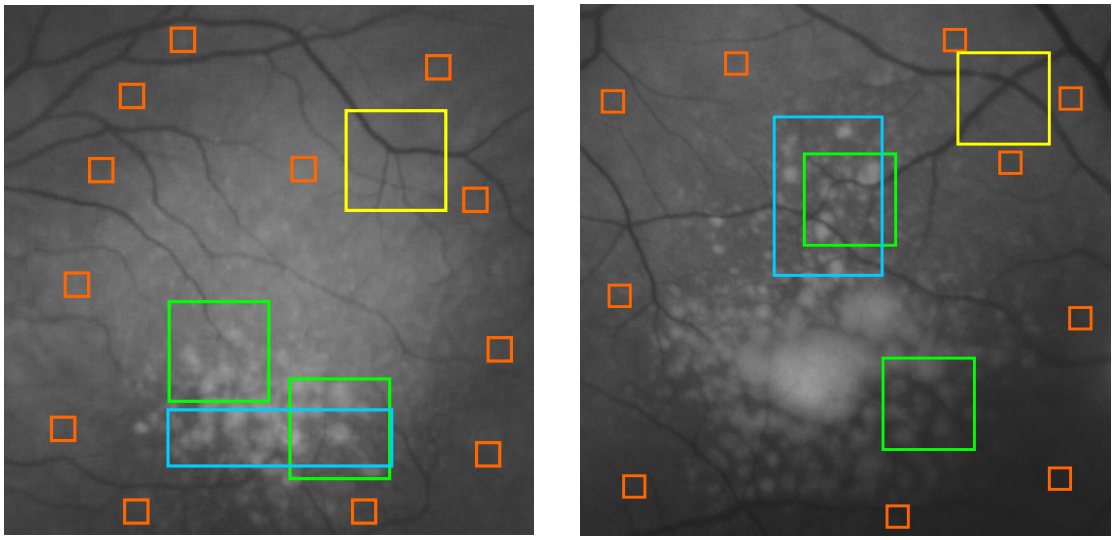


Figure 3.17 - Measurements regions for quantitative algorithms comparison.
Orange – Intensity; *Green* and *Yellow*– Contrast; *Blue* – Morphology.

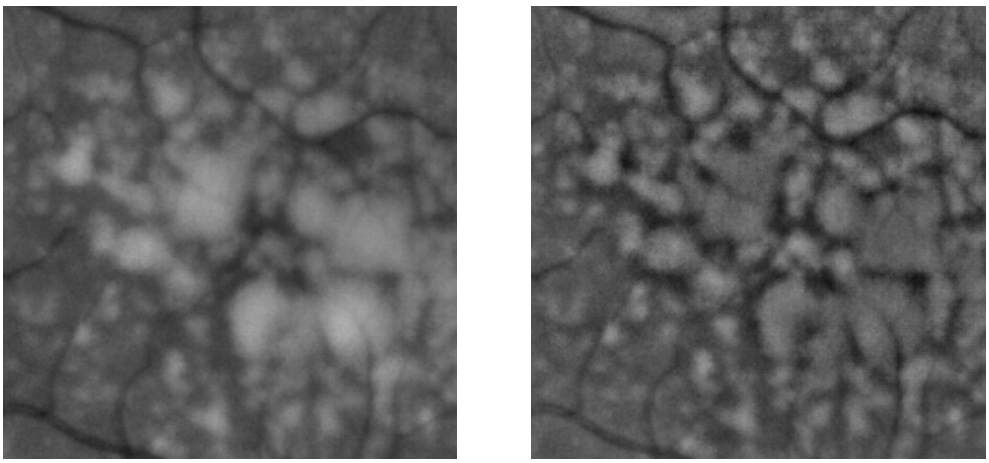
deviation of these latter values is the uniformity of intensity. A low value means a flatter background, while a higher value means that illumination varies along the image.

A good contrast between drusen and background will benefit the performance of the detection methods. However, a good contrast can be masked by positive and negative contrasts, which affect the drusen morphology, as it can be seen in figure 3.18. This is an unwanted effect that should be measured and used as a performance indicator. Therefore, two measurements were introduced: the RMS contrast and the correlation between the original image and the corrected one, named as contrast and morphology parameters, respectively.

The procedure to calculate the *contrast* parameter is based on the manual selection of two square regions containing drusen and background (*green* squares on figure 3.17), on the selection of one reference square containing vessels over background (*yellow* squares on figure 3.17), on the calculation of the RMS contrast (Peli 1990) on each of them (3.9), on the contrast normalization on the drusen areas by dividing their RMS contrast by the reference RMS contrast value and finally on the averaging of the two values.

$$RMS_{contrast} = \sqrt{\frac{1}{M \cdot N} \cdot \sum_{i=0}^{M} \sum_{j=0}^{N} \left(I_{ij} - \bar{I}\right)^2} \tag{3.9}$$

The *morphology* parameter measures the similarity on drusen areas, between the original image and the corrected one. The comparison should be done on areas where the original image contains drusen over an uniform background, to avoid its influence in the calculation. The procedure is to manually select one rectangular area (*blue* squares on figure 3.17) and calculate the Pearson Correlation Coefficient (Rodgers and Nicewander 1988) between the original image and each corrected one. In a correlation operation the addictive and multiplicative factors, such as the mean illumination, do not influence the result and only



(a) *(b)*

Figure 3.18 – Morphology changes after illumination correction. (*a*) original; (*b*) corrected.

the non uniform changes in the contrast are evaluated. The correlation operation returns a value inside the interval [0..1]. Zero if the drusen morphology differs from the original and one in an optimal solution when the two images are equally contrasted.

The overall evaluation of the algorithms will be a weighed sum of all three parameters. However, a normalization of the parameters is required to obtain a similar behaviour from each of them. The illumination parameter is lower on the better corrected images, while the contrast and the morphology are higher on these latter. Also, the parameters value range should be normalized to the interval [0..1]. Thus, it was assumed that the better corrected images should have the lower Cartesian distance from the origin (zero), i.e., having the lower combined value of all three parameters. The illumination was normalized with its division by its maximum value among the several techniques, the contrast was divided by its maximum and then inversed, and the morphology was inversed. Although the parameters are now normalized to [0..1], weighs (W_{illum}, W_{contr}, W_{morpho},) were introduced to rank their importance. These weighs were chosen empirically based on the observed values and on the importance of each parameter. The final evaluation is the Cartesian distance of all three parameters as in equation (3.10).

$$r_i = \sqrt{\left(w_{illum} \cdot \frac{I_i}{\max(I)}\right)^2 + \left(w_{contr} \cdot \frac{\max(C)}{C_i}\right)^2 + \left(w_{morpho} \cdot \frac{1}{M_i}\right)^2} \qquad (3.10)$$

$$\text{with weighs} \begin{cases} w_{illum} = 1 \\ w_{contr} = 2 \\ w_{morpho} = 10 \end{cases}$$

The comparison between the six algorithms is shown in figure 3.19 and figure 3.20. The first figure shows a graphical representation of the parameters, already multiplied by their

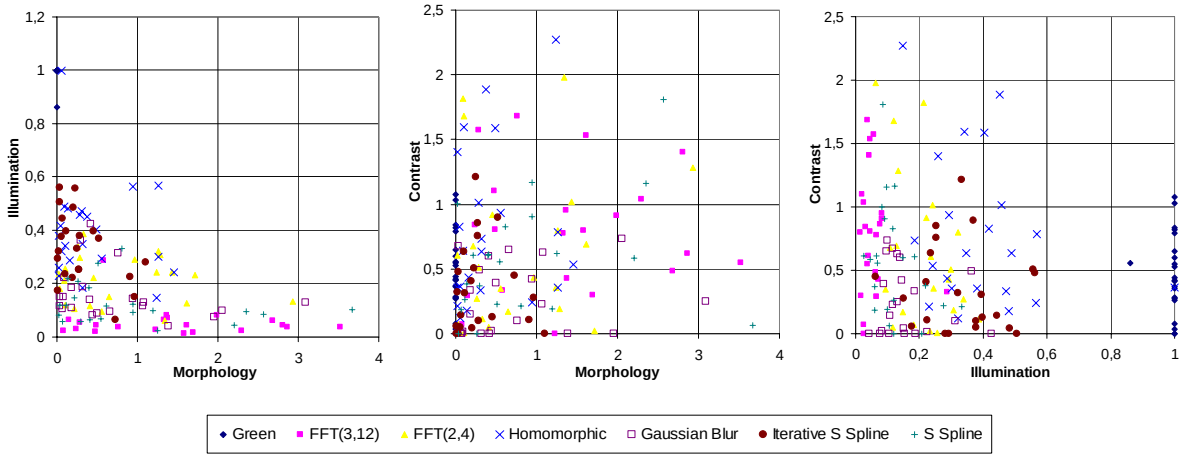
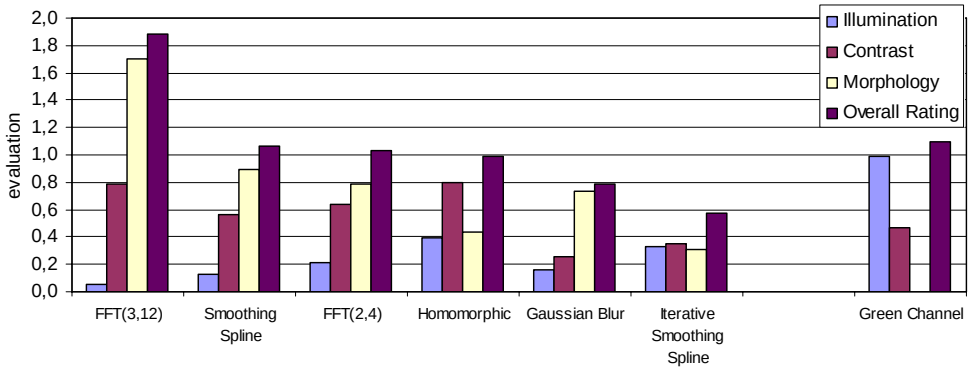


Figure 3.19 - Parameters graphical representation (with weighs).



Method	*Illumination*	*Contrast*	*Morphology*	*Overall Evaluation*
FFT(3,12)	0,06	0,79	1,71	1,88
Smoothing Spline	0,12	0,56	0,89	1,06
FFT(2,4)	0,21	0,64	0,79	1,03
Homomorphic	0,39	0,79	0,43	0,98
Gaussian Blur	0,16	0,26	0,73	0,79
Iterative Smoothing Spline	0,33	0,35	0,31	0,57
Green Channel	0,99	0,47	0,00	1,10

Figure 3.20 – Overall ranking of the correction algorithms.

weighs, for the 22 images. The second figure presents the overall evaluations for the several algorithms in graphical and table formats.

From this analysis it can be seen that the Iterative Smoothing Spline algorithm presents the lowest distance and therefore the best evaluation. From the plot in figure 3.20 it can be observed that it has the three parameters equally balanced and has the best morphology parameter, which is an important indicator. The illumination uniformity was not the most effective, but it is a compromise with the morphology parameter. An algorithm which only flattens the illumination behaves similarly with drusen areas and decreases the morphology parameter. Therefore, the ideal is an algorithm which flattens the illumination without losing the contrast and the morphology in the drusen areas, i.e., to have them balanced.

3.2.4. Contrast Normalization

The image obtained after the correction of the illumination non-uniformity has an average intensity value of 1 (one). To normalize it to a mid-scale value it might generate out-of-scale values as consequence of high contrasted images. Therefore, to facilitate further processing and provide images with full-scale values, the image intensity and contrast settings will be corrected to predefined levels.

To recreate a grey scale image with 256 levels, a constant value is multiplied by the image as if a constant light pattern was illuminating the retina. With the objective of having a higher contrast on the brighter areas the image is multiplied by 85 (1/3 of the grey scale). Such constant value was obtained empirically, in order to leave 2/3 of the grey scale for

representing the bright areas. Even though, depending on the image contrast, this value can generate underexposed or overexposed images. To overcome this problem a contrast normalization procedure was introduced. The procedure starts by evaluating the image contrast and then corrects it to a predefined value.

To quantify the contrast, after correcting the illumination, the image intensity standard deviation (*SD*), was measured. Theoretically, a high contrasted image will have a high *SD*, while a low contrasted image will have a low *SD*. Although the *SD* is influenced by retinal structures and irregularities, the overall image *SD* should not be used.

Hoover and Goldbaum (2001) used the retinal vessels intensity as the reference for evaluating and compensating the non-uniform illumination. The retinal vessels are one of the structures that are present in all images and providing good contrast with the background.

For this reason, it is herein proposed a method that evaluates the contrast between the vessels and the background and uses it to normalize the overall contrast. It is based on a sliding window analysis method which identifies vessels over background and determines the *SD* locally. The sliding window should be small enough to contain a vessel segment surrounded by background, in order to evaluate only their contrast. Considering that the vessels' diameter within the macular region is less than 90μm, a width of 1/8 of the optic disc diameter (≈ 180μm) was determined empirically.

In this analysis will appear windows containing only background, drusen or a combination of both with or without vessels. Therefore, it is necessary to evaluate which of

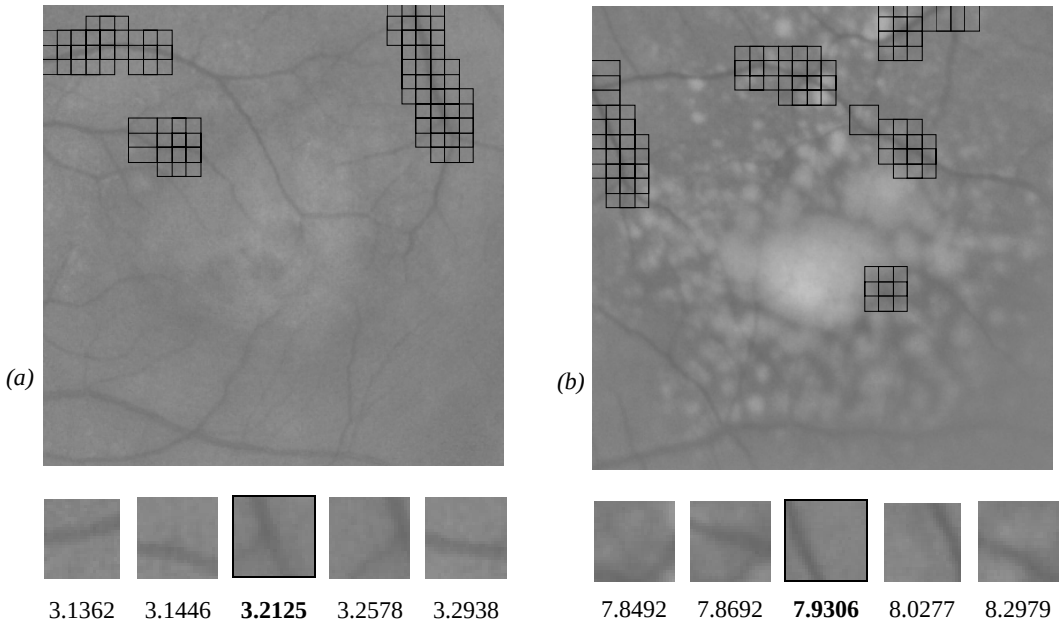


Figure 3.21 – Examples of the normalization procedure and their sliding windows with standard deviations (median is in bold).

these are candidates to be a reference *SD* window. As the vessels are the darker structures in the retina, the darker images should contain vessels and background. This observation motivated the use of the window mean intensity to select darker images. Then, from a selection of the darker windows (50 windows), the window containing the median of the *SD* observations is used as the contrast reference for the adjustment of the overall image contrast. This overall contrast was defined to be *norm_sd* = 15, a value chosen empirically with enough contrast without, however, producing over scale values.

As the automatic algorithms can fail in some particular situations (images defects or other retina irregularities), the selected window requires validation by the user. If it is not accepted, a correct reference window must be selected.

The procedure is as follows:

1. The image is analyzed in a sliding window scheme and for each window the *SD* and the mean intensity are calculated;
2. From the 50 darker windows, that are typically windows containing vessels over a uniform background, is determined the median *SD* (*median_sd*) (figure 3.21).
3. The median *SD* sliding window is shown to the user to validate the automatic selection. If the window doesn't contain vessels over a flat background, the user must select a correct window and the *median_sd* is updated to this new window.
4. The image intensity is updated using the equation below, where *norm_sd* is the predefined standard deviation (*norm_sd* = 15) and *A* the constant background value (*A* = 85):

$$norm\_image(x,y) = A \cdot \left(\frac{norm\_sd}{median\_sd} * (image(x,y) - 1) + 1 \right) \qquad (3.11)$$

Figure 3.22 shows two images before and after the intensity and contrast normalization. In these test images the improvement on contrast is clearly visible, maintaining the same levels in both images. The images are now ready for the automatic drusen detection.

3.3 - Drusen Detection

The drusen detection is an intermediate step in the drusen quantification. The modelling of drusen spots which will be described in section 3.4 -requires the initialization of the model parameters in order to reduce the search space for the fitting algorithm. The drusen, the only structures modelled from the retina, are initially characterized by the location and

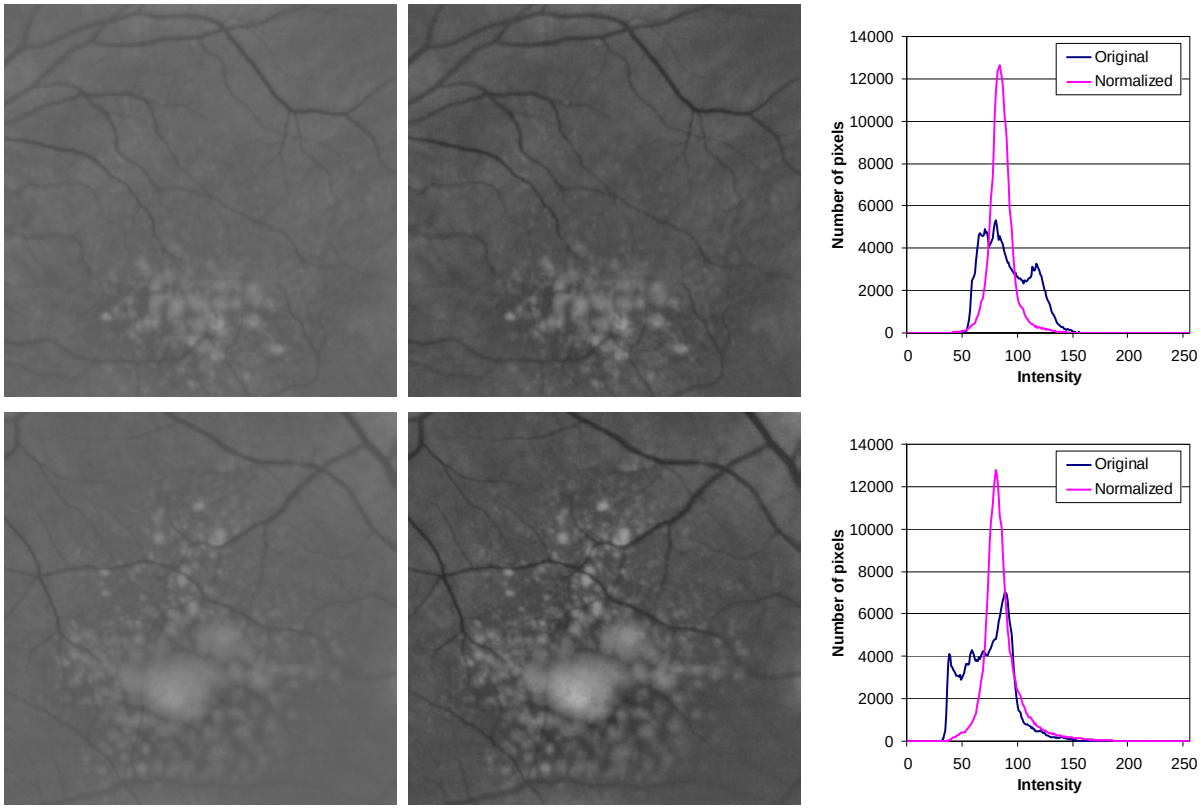


Figure 3.22 – Images and their correspondent histograms before and after intensity and contrast normalization..

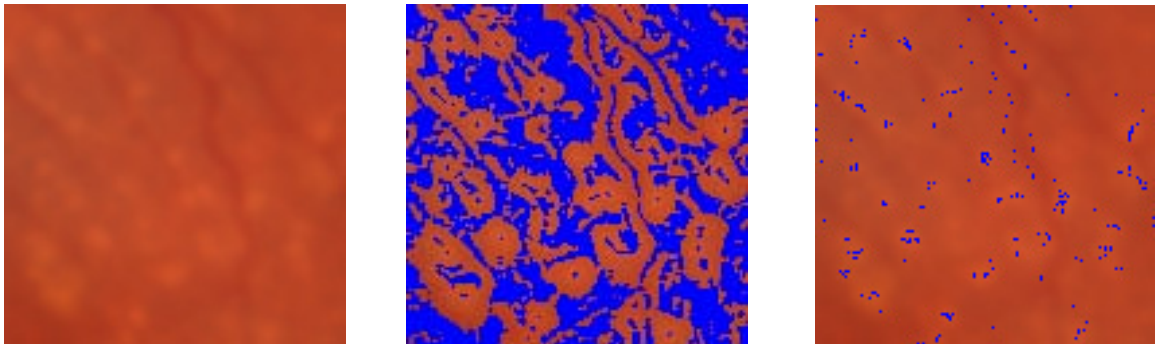
peak value amplitude. This information is provided to the modelling algorithm which will be presented in the next section.

When drusen are isolated from the others, they have an approximately round shape and their maximum amplitude located in the centre. However, in confluent drusen they are superimposed to other drusen and their round shape is added to the others, assuming a different overall shape. These superimposed drusen should also be characterized in order to have a detailed model. In these cases their location and amplitude will be more difficult to determine.

The image noise is another obstacle to the reliability of the local maximums detection algorithms. The irregularities it provokes are easily misclassified as maximum intensity pixels, generating over-segmented images. In this work it was developed an algorithm that can reduce the noise influence and provide additional spot characterization, which is used to reject drusen unrelated spots.

3.3.1. Regional Intensity Maximums Detection Algorithms

One of the most common algorithms for regional intensity maximums detection in signal processing is based on the analysis of the first derivative of the signal. The zeros of the first derivative are plateaus or are inflection points which can be maximums or minimums. In



(a) *(b)* *(c)*

Figure 3.23 – Regional Intensity Maximums detection methods. (a) original image; (b) Threshold over the Gradient (approximate zero); (c) Threshold over the Laplacian filter.

image processing the first derivative operation is obtained by gradient operators such as Roberts, Prewitt or Sobel edge detection filters. By applying a threshold to their output, in order to select the pixels with a gradient approximately of zero, it will generate a high number of candidate pixels. In figure 3.23 (a,b) is showed an example of the candidate pixels (in blue) obtained using the Sobel operator.

The second derivative by means of the Laplacian filter is also commonly used as a blob / point detector (Rosenfeld and Kak 1982). The filter uses a mask which intensifies the pixels containing lower neighbours (figure 3.24). By applying the Laplacian filter and thresholding the resultant image, it can be obtained an image with less candidate pixels than the previous method. Figure 3.23.c presents the final output from this detection method.

-1	-1	-1
-1	8	-1
-1	-1	-1

Figure 3.24 – Laplacian filter mask.

However, both methods are dependent on adequate threshold selection to obtain acceptable results and have high tendency to generate over-segmented images. For these reasons they were not considered suitable for regional intensity maximums detection. A more reliable algorithm, the Watershed Transform, was then tested for this purpose.

3.3.2. The Watershed Transform

Beucher and Lantuejoul (1979) were the first to use the concept of Watersheds for image segmentation. They used it for bubbles detection in radiographic plate and to display facets in a metallic fracture. The algorithm uses the principle of geographical watersheds, where during a rainfall the water will flow into a *catchment basin*, areas draining into a river

or reservoir. The *watersheds* are the lines dividing the domains of attraction of the rain (figure 3.25). An alternative approach is to imagine a landscape being immersed with water starting its local *minima's*; when the water from two *catchment basins* meet, those places are marked as *watersheds*.

Seeing a grey-scale image as a topographical image where the image intensity represents the height, and the *watershed* is a line of pixels separating two *catchment basins*. These *watershed* pixels define the limits for the segmented region, in this case a segmentation based on the pixel's intensity. The regional intensity maximums are obtained by computing the negative of the image, by applying the Watershed Transform and by determining location

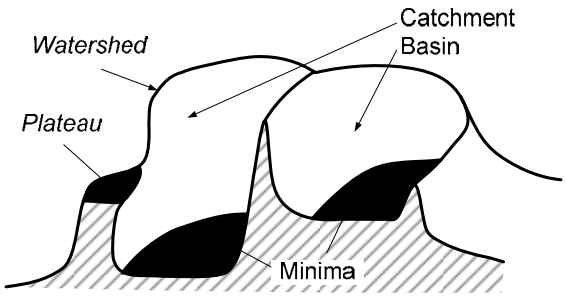
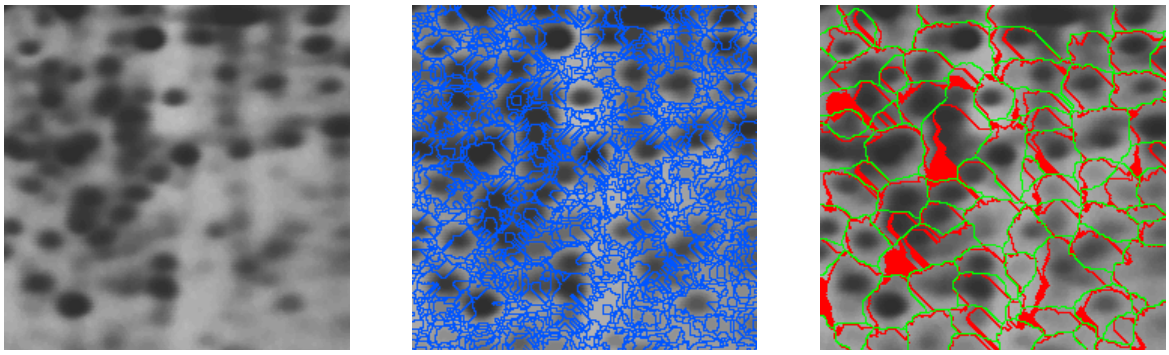


Figure 3.25 – Geographical Watersheds.

of the minimum intensity within each *catchment basin*.

The Watershed Transform in its original form (Beucher and Lantuejoul 1979) produces an over-segmentation of the image due to the existence of many local *minima*, which is undesirable (figure 3.26.b). One of the main causes is the presence of *plateaus*; here the water does not flow, leading to the creation of multiple labels. Several other approaches have been implemented in order to reduce this effect. They can be divided into two classes, one based on the immersion algorithm by Vincent and Soille (1991) and other based on topographical distance by Meyer and Beucher (1990). From the analysis of figure 3.26.c,



(a) *(b)* *(c)*

Figure 3.26 – The Watershed Transform. (a) Gel image (original); (b) by Beucher and Lantuejoul (blue *watershed*); (c) by Vincent and Soille (red *watershed*) and by Meyer and Beucher (green *watershed*). (*image courtesy of Dr. S. Beucher, CMM/Ecole des Mines de Paris*).

which shows the results from the two approaches, they produce the same number of labels and only on *plateaus* their differences are visible.

The two classes differ basically in the way the *watersheds* are defined in some *plateaus* and in some crests. A good description on watershed definitions and implementations can be found in Roerdink and Meijster (2000) article. Here are discussed the different mathematical definitions for the Watershed Transform and presented sequential implementations strategies, as well as strategies for the parallelization of the algorithm.

The Vincent and Soille Immersion algorithm (1991) is from the class of the flooding methods, considered the fastest and most accurate algorithms, to compute watersheds (Hagyard, Razaz et al. 1996). It operates in two stages: a *sorting* step followed by a *flooding* step to propagate *labels* from the image minima to the topographical crests. It uses a *label* image, where *labels* and *watershed* are defined, and a work image, to store geodesic *distances*. In the *sorting* step, the pixels are stored in an ordered list by their grey scale intensity. Then in the *flooding* step the pixels are analyzed at each intensity level in turn. Intensity *minima* receive unique *labels* which are then propagated using a breadth-first implementation as follows. The pixels at level *h* that have lower intensity neighbours are labelled with its neighbour's *label*; are added to a first-in first-out queue (*FIFO*); and receive the geodesic distance of their neighbour's plus 1. These pixels are then removed from the queue and their unmarked neighbours from the same intensity level *h*, are assigned with the same *label* and placed on the queue. This allows spreading the *labels* within the same intensity level, especially in the *plateaus*. The *watershed* pixels are assigned to the pixels that have different neighbour *labels*, including the spreading of *watersheds* within *plateaus*. This wields to the creation of thick *watersheds*, which need to be corrected using alternative methods based on the topographical distances or are maintained as it is (see figure 3.26.c). This flooding method could be seen as a top-down approach, where the pixels of higher altitude get their *label* from the lower intensity neighbours.

The authors Meyer and Beucher (1990) proposed to solve the over-segmentation and the thick *watershed* lines provoked by the *plateaus* with an implementation of the Watershed Transform by means of a topographical distance. To propagate the *labels* progressively along the *plateaus*, from the lower to the upper boundary, they transformed the image to be *lower complete*, that is, "*each pixel which is not in a minimum has a neighbour of lower grey value*" (Meyer and Beucher 1990).

When filling a valley containing non-minima *plateaus* with water, it will be progressively propagated from the lower boundaries to the upper boundaries of the *plateau*,

and not instantaneously. This physical principle is the basis for the *lower-complete* operation. It transforms the *plateaus* into slopes using the pixel's geodesic distance to the *plateau*'s lower boundaries. The method detects the *plateaus* and increases the intensity of its pixels according to their distance to the lower boundary.

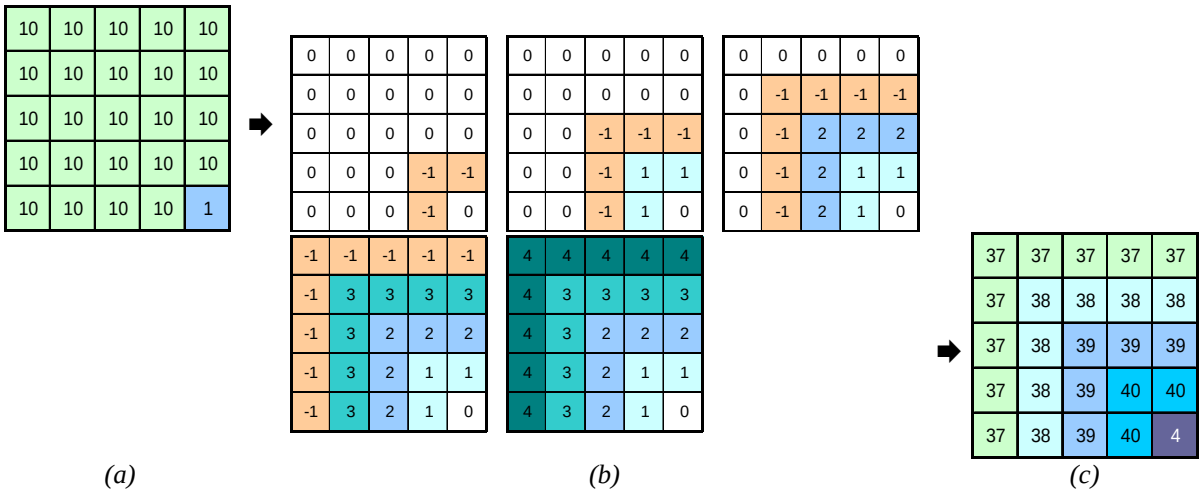
The *lower-complete* transform as proposed by Meyer and Beucher (1990) is a mapping between the image $f(x,y)$ and the *lower-complete* image $l(x,y)$, defined as:

$$lc(x,y) = \lambda \times f(x,y) + d(x,y) \text{ where } \lambda > d(x,y), \ \forall x,y \tag{3.12}$$

The image $d(x,y)$ contains, for each pixel inside the *plateaus*, the minimum geodesic distance to the *plateau* lower boundary. To compute the distance image $d(x,y)$ an ordered propagation of *labels* is performed using a breath-first search (Cormen, Leiserson et al. 2001). The method uses an ordered FIFO queue to evaluate progressively the distance inside the *plateaus*. However, this method is adapted to flood from *minima* and doesn't solve *minima plateaus* (*plateaus* without lower neighbours), maintaining them flat.

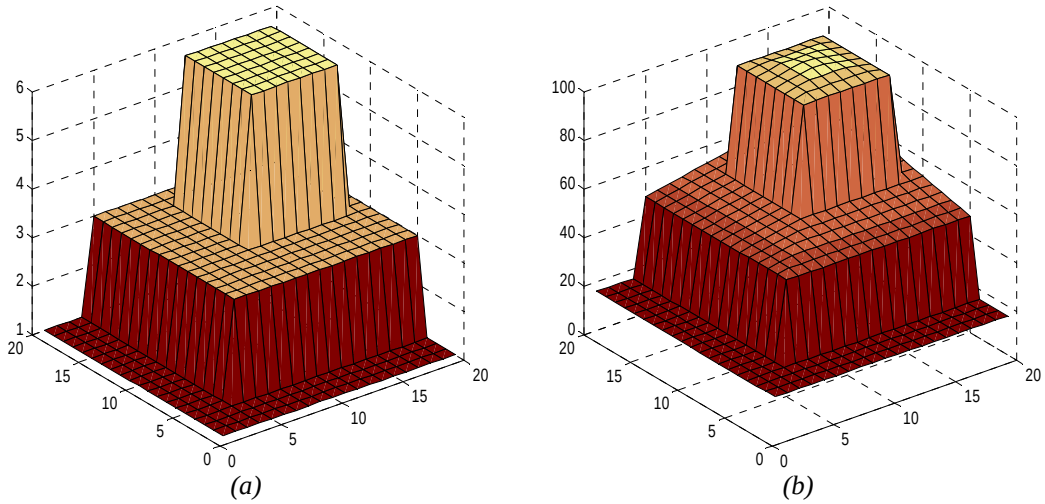
The algorithm to create *lower-complete* images (described in Algorithm 3.2) starts by assigning a special *label* (*MARKER*) to the pixels which contain lower intensity neighbours and adds them to a FIFO list (figure 3.27). These will be the seeds for the propagation algorithm, which follows the breath-first search principle.

On a second stage, these pixels will be taken out from the FIFO queue and will receive a unitary distance. Simultaneously, each of their neighbours' pixels, which have an equal intensity and are not yet labelled, will be marked and added to a FIFO queue containing the next distance level. After processing all the *labels* from the FIFO queue, the distance is increased by one unit and the process is repeated with the newly created queue. This



(a) *(b)* *(c)*

Figure 3.27 – Lower complete image generation. (a) original image ($f(x,y)$); (b) breath-first distance propagation from top-left to bottom-right direction ($d(x,y)$); (c) *lower-complete* image ($hc(x,y)$).



(a) *(b)*

Figure 3.28 – Lower complete diagram. (a) Original image; (b) *lower-complete* image.

propagation through the neighbours guarantees the sequential distance assignment to the pixels within a *plateau* using the breath-first search principle. The figure 3.27.b shows the successive iterations of the distance propagation for an image containing one *plateau*.

The distance image $d(x,y)$ generated after processing all the FIFO queues is finally used to compute the *lower-complete* image using a similar equation to Meyer and Beucher's *lower-complete:*

$$lc(x,y) = \begin{cases} \lambda \times f(x,y) + d(x,y) - 1 & \text{if} \quad d(x,y) \neq 0 \\ \lambda \times f(x,y) & \text{else} \end{cases} \tag{3.13}$$

where:

$$\lambda = \max(d(x,y)) + 1 \tag{3.14}$$

The result is an image without *plateaus* except those on regional *minima* (figure 3.27.c). It is overscaled in order to support the creation of a slope within the *plateaus* while maintaining the proportions between intensities. In figure 3.28 it is presented a three-dimensional view of an image containing non-maximum *plateaus* and its *lower-complete* image, where the substitution of the *plateaus* by slopes is clearly visible.

It should be noticed that in the algorithm 3.2 there is only one FIFO queue which will receive all the distance levels. A special *label* (FIFO_END) is added to the queue working as a separator between them. When the iterative process reads a FIFO_END the distance is advanced by one unit.

Algorithm 3.2 - Lower-Complete Image
1: *lower_complete* = **Function** *Lower_Complete_Image*(*image*) **{**
2:
3: *// For all the image pixels:*
4: *// - Initialize queue with pixels that have a lower neighbour*
5: *// - set others to 0*

```
6:      foreach( p in image ) {
7:          lower_complete( p ) = 0;
8:
9:      // if p has a lower neighbour then add it to the FIFO
10:         if (has_lower_neighbour( p )) {
11:             fifo_queue->Add( p );
12:
13:             lower_complete( p ) = MARKER;
14:         }
15:     }
16:
17:     int dist = 1;
18:
19:     fifo_queue->Add( FIFO_END );
20:
21:     while (fifo_queue->Count != 0) {
22:         p = fifo_queue->RemoveFirst( );
23:
24:         if (p == FIFO_END ) {
25:             if (fifo_queue->Count != 0) {
26:                 fifo_queue->Add( FIFO_END );
27:
28:                 dist = dist + 1;
29:             }
30:         }
31:         else {
32:             lower_complete( p ) = dist;
33:
34:           // for all q belonging to p neighbourhood with (im[q] = im and lc[q] = 0)
35:           // add it to the queue
36:             foreach( q in neighbours( p ) ) {
37:                 if ((image( p ) == image( q )) and (lower_complete( q ) == 0)) {
38:                     fifo_queue->Add( q );
39:
40:                     lower_complete( q ) = MARKER;
41:                 }
42:             }
43:         }
44:     }
45:     //Put the lower complete values in the output image
46:     foreach( p in image ) {
47:         if (lower_complete( p ) != 0)
48:             lower_complete( p ) = dist * image( p ) + lower_complete( p ) - 1;
49:         else
50:             lower_complete( p ) = dist * image( p );
51:     }
52: }
```

The next stage of this Watershed Transform method is based on an ordered flooding of *labels*. The pixels are labelled following the shortest path to the *minima*. It is a recursive method that uses an ordered queue, which is an array of *h FIFO* queues, one for each grey level in the image. The ordered queue is initialized with the markers from where the flooding should start, these could be the *minima* intensity or user specified pixels. A pixel of *h* grey level will be placed on the h^{th} *FIFO* queue with a unique *label*. The queue is then processed from the lowest grey level to the highest. The pixels are removed from the *FIFO*, one at a time, and their *label* is assigned to their non-labelled neighbours of higher or equal intensity, which are then inserted into the *FIFO* queue in their corresponding *h* grey level. Once an *h* grey level *FIFO* queue is empty, the algorithm moves to the next higher grey level. The

procedure finishes when the *FIFO* queue is totally empty. The *labels* using this method are propagated in the *plateaus* in a synchronized breath-first order coming from the different *minima*.

This Watershed Transform method, in opposition to the previous method, ensures a bottom-up propagation of *labels*, which are propagated from the lower intensity to the higher intensity pixels. Although, the drawbacks from this method are that, it requires the initial *labels* to be defined and that it does not produce *watershed* pixels, requiring a post-processing to assign *watershed* pixels to the *labels* boundaries (figure 3.26.c).

To implement the current method, several methods were proposed by the authors Meyer and Beucher (1990; 1993; Meyer 1994) namely, ordered scanning based on integration via the Dijkstra-Moore shortest paths algorithm, ordered scanning by Hill Climbing algorithm or sequential scanning based on image integration algorithm. From these three the Hill Climbing method is the less complex and more efficient, as it does not require the distance function to be computed, nor the ordered queues to be implemented. The pseudo-code for this method is presented in algorithm 3.3

Algorithm 3.3 – Watershed Transform by Hill Climbing

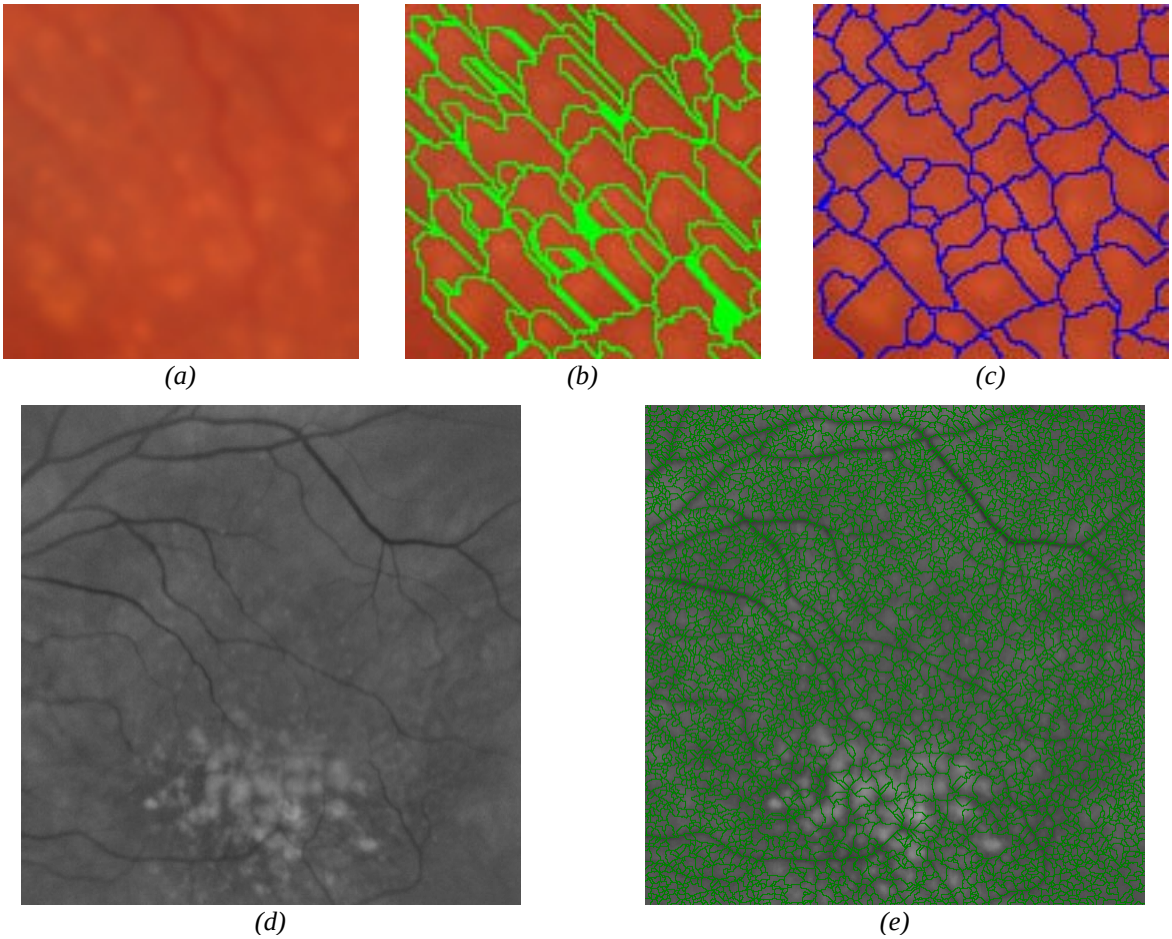
```
1:  label_image = Function Watershed_Hill_Climbing(image) {
2:
3:      #define WSHED 0
4:      #define MASK -1
5:
6:  // Get Image Lower Complete
7:      LC_Image = Lower_Complete_Image (image);
8:
9:  // Initialize image lab with distinct labels for minima and special label mask for all other pixels
10:     Label_Init(image, label_image);
11:
12: // Fill the H_fifo_queue with Minima boundary pixels
13:     H_fifo_queue = Get_Initial_Seed_Pixels(image, label_image);
14:
15:     int currentInt = 0;
16:     while (currentInt < H_fifo_queue->Length){
17:         if (is_not_empty(H_fifo_queue[currentInt]g) )    {
18:         //  take pixel with minimal grey value from queue
19:             p = fifo_queue[currentInt]->RemoveFirst();
20:
21:         // label steepest upper neighbours
22:             foreach( q in neighbours( p ) ) {
23:                 if (image (p) < image(q)) {
24:                     if (label_image (q) == MASK) {
25:                         label_image(q) = label_image(p);
26:                     }
27:                 }
28:             }
29:         }
30:     // process next H_level
31:         currentInt = currentInt + 1;
32:     }
33: }
```

The Watershed Transform applied to retinal images (figure 3.29) produces acceptable results for the detection of maximums. It detects the majority of the regional intensity maximums, but tends to produce over-segmented images. The execution time, although is linear with the number of pixels, is long as consequence of neighbouring analysis and of lists management.

3.3.3. Gradient Path Labelling

The detection method that is proposed in this thesis is a novel segmentation algorithm which uses the image gradient as the basis of a pixels' labelling procedure. It is a maximum intensity detection algorithm based on the labelling of gradient paths.

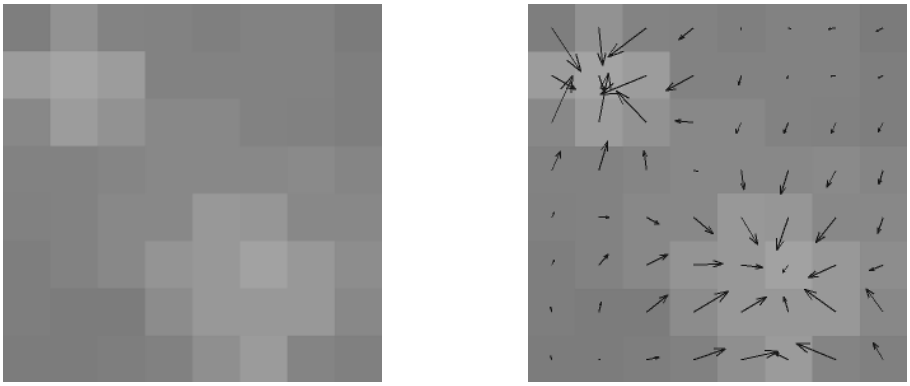
By looking at an image gradient in a *quiver plot*[1] it is noticed that in the drusen proximity the image gradient vectors are in the direction of an intensity maximum. This is visible in figure 3.30.(a)(b) where a zoom on the gradient of a two small drusen image with 10x10 pixels, is represented. Similarly to a mountain climb, there are one or more paths that reach the same maximum intensity. The confluence of several significant ascending paths



(a) *(b)* *(c)*

(d) *(e)*

Figure 3.29.- Watershed Transformed applied to retinal images. (a,d) original image; (b) Vincent and Soille method; (c,e) Meyer and Beucher method.



(a) *(b)*

Figure 3.30 – Image gradient. (*a*) Original image (zoom); (*b*) Image and gradient quiver plot.

means that we are in presence of a regional intensity maximum (RIM). This is the principle of the *Gradient Path Labelling* algorithm (GPL).

The algorithm detects RIM and also defines the drusen areas of influence, composed by the pixels belonging to the drusen ascending paths. The proposed algorithm groups ascending paths that belong to the same RIM. The result is a segmented image, where drusen spots are isolated from each other.

The implementation has some similarities to the *connected components* algorithm (described in (Gonzalez and Woods 2007)) where the labels are propagated to the adjacent pixels if they verify a merging condition. These conditions are usually based on pixels' intensity or colour similarity. In the *connected components* algorithm one of the objectives is to isolate objects from each other and from the background. As for the GPL algorithm the purpose is to separate objects with their surrounding background. The output from the GPL algorithm is similar to Beucher's and Lantuejoul's Watershed Transform (1979), although it is faster and is optimized to maximums search.

The method is divided into two main stages, *labelling* and *labels merging*, which are described in the following sections.

3.3.3.1 Labelling

The *Labelling* of an image is a procedure designed to group pixels with similar features, and usually adjacent pixels, by assigning them common *labels*. In this work the objective is to separate drusen spots by identifying the pixels belonging to them. The adopted procedure is to group pixels which belong to ascending intensity paths that finish on the same RIM, i.e., the same druse spot. To evaluate the ascending paths the GPL labelling procedure uses the gradient image.

The gradient image is determined by computing first derivative of the image intensity. The most common methods in image processing to perform this operation are the edge detection methods, such as Roberts, Prewitt or Sobel filters (Duda and Hart 1973). The main advantage of this gradient approach is the analysis of pixel's 3x3 neighbourhoods that eliminate several noise related artefacts.

The gradient of a two dimensional image, $f(x,y)$, is defined by the vector:

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{3.15}$$

As it can be seen, the image gradient evaluates the derivatives in horizontal (G_x) and vertical (G_y) directions and combines them. The angle of the vector is very important here, as it indicates the direction to the maximum rate of change in $f(x,y)$, i.e., the direction to the ascending intensity path. The vector modulus will determine the strength of the change; however it is not used in our model.

The modulus and angle of the gradient vector can be approximated by the equations:

$$\begin{aligned} |\nabla f| &= |G_x| + |G_y| \\ angle(\nabla f) &= \tan^{-1}\left(\frac{G_y}{G_x}\right) \end{aligned} \tag{3.16}$$

For applying it to an image the G_x and G_y are evaluated in each pixel's neighbourhood, which is the procedure used in the edge detection methods enumerated before. The simplest and fastest is the Roberts filter which evaluates both directions in a 2x2 neighbourhood, although it is less accurate. The Prewitt and Sobel use the 3x3 pixel's neighbourhoods, but with different filter masks which are shown on figure 3.31. For these filters the horizontal G_x and vertical G_y derivatives are obtained by the product of the pixel neighbourhood by the two filter masks. Although the Prewitt filter is slightly simpler to implement (less four addition operations) it tends to produce somehow noisier results (Gonzalez and Woods 2007). This observation justified the use of the Sobel filter masks as the horizontal and vertical gradient kernels. The result is shown in figure 3.30.b.

p_1	p_2	p_3
p_4	p_5	p_6
p_7	p_8	p_9

Pixel neighbourhood

-1	0	1
-1	0	1
-1	0	1

-1	-1	-1
0	0	0
1	1	1

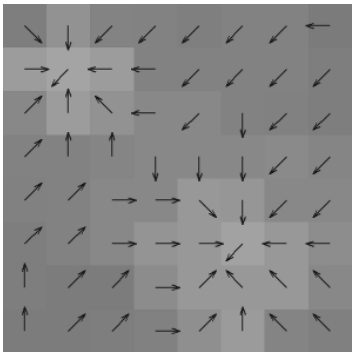
Prewitt

-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

Sobel

Figure 3.31 – Horizontal and vertical filter masks for Prewitt and Sobel edge detection filters, with respect to the centre point of the neighbourhood.



(a)

1	2	3	4	5	6	7	8
9	1	4	5	6	7	10	11
1	12	6	6	7	10	11	13
14	15	16	7	17	10	13	18
19	20	21	7	7	10	18	22
23	24	25	25	25	7	22	26
27	28	29	30	7	31	32	33
34	35	36	37	37	38	39	40

(b)

Labels		
1	⇔	-
2	⇔	1
3	⇔	1
4	⇔	1
5	⇔	4
6	⇔	1
7	⇔	-
8	⇔	7
9	⇔	1
10	⇔	7
11	⇔	10
12	⇔	1
13	⇔	10
14	⇔	12
15	⇔	12
16	⇔	6
17	⇔	7
18	⇔	7
19	⇔	15
20	⇔	16
21	⇔	7
22	⇔	7
23	⇔	20
24	⇔	21
25	⇔	7
26	⇔	22
27	⇔	23
28	⇔	25
29	⇔	25
30	⇔	7
31	⇔	25
32	⇔	7
33	⇔	22
34	⇔	27
35	⇔	29
36	⇔	30
37	⇔	31
38	⇔	31
39	⇔	31
40	⇔	32

(c)

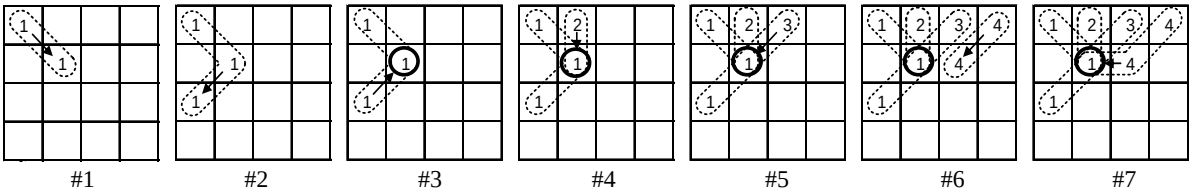
Figure 3.32 – GPL Labelling procedure. (*a*) Pixel search directions; (*b*) *Initial label propagation* (with label propagation on the two upper rows); (*c*) Labels *equivalence* table.

For assessing the pixels' adjacency in the spots detection it was assumed that horizontal, vertical and diagonal connectivity were allowed, i.e., *8-connectivity*. Therefore, to follow an ascending path the gradient angle should be rounded to one of the possible eight search directions (N, NE, E, SE, S, SW, W, NW). The image gradient from figure 3.30 should then look like the one in figure 3.32.a after this round operation.

The second stage of the labelling procedure, *initial label propagation*, examines every pixel in a top-left to bottom-right direction and assigns a new label to the pixels that are not marked. The *label* is then propagated to the pixel pointed out by the gradient angle and will continue to be propagated until the next one is already marked or is outside the image boundaries. Figure 3.33 shows an example of the label propagation for the first four *labels*. Here two of the stopping conditions are exemplified.

Every time the label propagation finishes on the same *label* that is being propagated, the pixel is marked as a RIM (*label 1* in figure 3.33) and its coordinates and amplitude are added to a list containing all RIM descriptions. Whenever it finishes on a different *label*, these two *labels* are defined as *equivalents*, i.e., they belong to the same RIM (*labels 2, 3 and 4* in figure 3.33), because their ascending paths will be coincident here on. These *labels equivalencies* are then stored in a list (figure 3.32.c) containing all the *equivalencies* to be further processed and grouped.

Other stopping conditions occur in the image boundaries when the propagating direction is beyond the image limits. To solve it, two solutions are suitable. One is to define



#1 #2 #3 #4 #5 #6 #7

Figure 3.33 – Example of the sequential *label* propagation procedure.

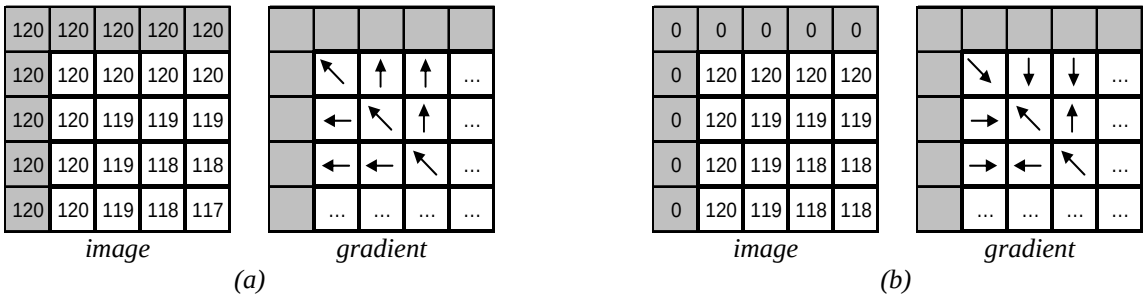


Figure 3.34 – Gradient calculus on image boundaries; (a) boundary replication; (b) 0's boundaries.

the pixels as RIM, since the direction is invalid, but it generates over-segmented regions. Another is to avoid the search direction that goes outside the image boundaries, which can be done by adding rows and columns of 0's surrounding the image when computing the image gradient. The result is a gradient image whose gradient directions never go outside the image boundaries (figure 3.34). This solution reduces the image over-segmentation without compromising the final segmentation.

After finishing the labelling process for all the pixels, the *label equivalences* are applied. The *equivalences' table*, as illustrated in figure 3.32.c, uses a simplified transitive closure procedure in order to find, for each *label*, its lowest equivalent *label*. These new *equivalences* will be propagated over image pixels generating an image containing a smaller number of *labels*. The result is an image with as many *labels* as RIM detected and where each RIM is characterized by its amplitude, location and area of influence, i.e. all the pixels which gradient path ends on that RIM (figure 3.35).

The *labelling* stage of the GPL method is described in the algorithm 3.4. The complexity and time consumption of this implementation is approximately linear to the size of the image, what is considered an advantage. Only the *equivalences table* processing is not linear to the number of *labels*, although its complexity is very low and not comparable to the remaining part of the algorithm, not influencing the overall algorithm performance. Note that so far it is a non-parametric RIM detection algorithm.

1	1	1	1	1	1	7	7
1	1	1	1	1	7	7	7
1	1	1	1	7	7	7	7
1	1	1	7	7	7	7	7
1	1	7	7	7	7	7	7
1	7	7	7	7	7	7	7
1	7	7	7	7	7	7	7
1	7	7	7	7	7	7	7

Figure 3.35 – GPL final labels image.

Algorithm 3.4 - **Gradient Path Labelling**

```
1:  label_image = Function Gradient_Path_Labelling(image) {
2:
3:      // Get gradient directions
4:          [gradient_x, gradient_y] = Sobel_filter (image);
5:          [delta_x, delta_y] = Round_8conn(angles_matrix);
6:
7:      // Initial configuration
8:          current_label = 1; // initial label
9:          label_image = matrix_zeros(sizeof(image)); // empty matrix for storing labels
10:
11:     // Raster iteration over all image pixels
12:         for(y = 0; y < img.height; y++) {
13:             for(x = 0; x < img.width; x++) {
14:
15:                 // execute only empty labels
16:                     if (label_image(x,y) == 0) {
17:                     // set new label
18:                         label_image(x,y) = current_label;
19:
20:                     // next pixel coordinates
21:                         next_x = x + delta_x(x,y);
22:                         next_y = y + delta_y(x,y);
23:
24:                     // label propagation
25:                         while ((is_inside_boundaries(next_x, next_y)  and
26:                                 (label_image(next_x, next_y) == 0)) {
27:
28:                             // set the label
29:                             label_image(next_x, next_y) = current_label;
30:
31:                             // advance to the next pixel
32:                             old_x = next_x;
33:                             old_y = next_y;
34:
35:                             // next pixel coordinates
36:                             next_x = old_x + delta_x(old_x,old_y);
37:                             next_y = old_y + delta_y(old_x,old_y);
38:                         }
39:
40:                     // stop condition: outside the image, regional maximum, merge labels
41:                         if ( not(is_inside_boundaries(next_x, next_y))
42:                         // outside the image boundaries
43:                             regional_maximums.add(next_x, next_y);
44:                         else if (label_image(next_x, next_y) == current_label)
45:                         // add a new regional maximum
46:                             regional_maximums.add(next_x, next_y);
47:                         else if (label_image(next_x, next_y) != current_label)
48:                         // define them as compatibles
49:                             equiv_tbl.add(current_label, label_image(next_x, next_y));
50:
51:                     // set next label value
52:                         current_label = current_label + 1;
53:                     }
54:             }
55:         }
56:
57:     // apply equivalences
58:         label_image = apply_equivalences(label_image, equiv_tbl);
59: }
```

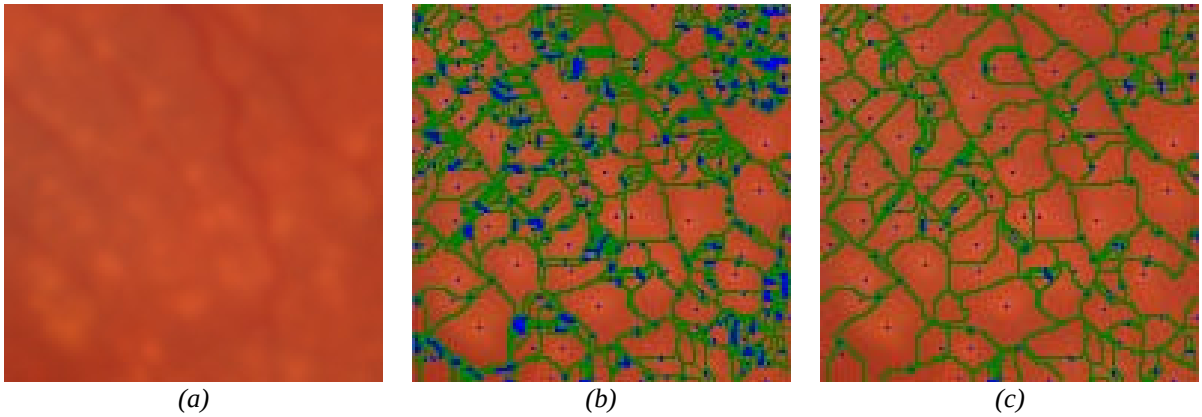
3.3.3.2 Labels Merging

After the labelling stage the image is segmented. However, when *plateaus* exist (flat valleys or flat hills) not all gradient paths end on the same RIM pixel, which generates an over-segmented image. Consequently, the algorithm will generate more RIM then it should, which are not relevant for the analysis being performed. An example is shown in figure 3.36 where image *(a)* was over-segmented by the GPL algorithm resulting on image *(b)* (the blue dots are the RIM and green pixels are the label boundaries).

Following the method used by Meyer and Beucher for solving the *plateaus* problem in the Watershed Transform it was applied to the retinal images to evaluate its effectiveness. In the retinal images there are more plateaus in the background rather than on drusen maximums. Therefore, the *lower-complete* transformation was adapted to transform the image to be *higher-complete* and used in the detection of regional maximums. Instead of maintaining the minima *plateaus*, in a *higher-complete* transformation it maintains the maximum *plateaus*.

By applying the *higher-complete* transformation (Meyer and Beucher 1990) to the image, before applying the gradient, the pixels within the *plateaus* will have gradient vectors directed to the regional maximums. The consequence is that the GPL will achieve a better performance, producing more accurate regions. In figure 3.36.c it is shown a segmented image with the use of the *higher-complete* transformation. Here the number of detected RIM decreased from 876 to 195, what can be considered a good achievement. However, although this was a good decrease, the over-segmentation still exists. To solve this problem a *merging algorithm* was adopted.

The *merging* algorithm is based on the analysis of the paths with higher intensity between a RIM and each of its neighbours. The merging condition is for the path's amplitude difference to not exceed a predefined value, *d*. This parameter defines the level of detail of the



(a) *(b)* *(c)*

Figure 3.36 – Over-segmentation provoked by *plateaus*: (a) original image; (b) segmented with GPL algorithm (876 RIM); (c) using the *higher-complete* transformation and the GPL algorithm (195 RIM).

analysis. A high value will decrease the level of detail by merging more RIM, while a low value will have a tendency to maintain the over-segmentation. It is calculated as a percentage of the RIM amplitude, having a fixed minimum difference value (equation (3.17)). In the equation the *max* and *min* operators relate to the maximum and minimum intensity with the pixels that contribute to the same RIM.

$$d = \begin{cases} \left(\max(RIM) - \min(RIM)\right) \cdot d_{\%} & if \quad > d_{fixed} \\ d_{fixed} & else \end{cases} \tag{3.17}$$

In this work, the choice for the *d* value was done empirically with a set of normalized images. It was found a percentage of $d_{\%}$= *10 %*, with a minimum value of d_{fixed} = 3.

In figure 3.37.a is a two-dimensional example in which for two drusen, four corresponding RIM were detected. The maximums M_1, M_2 and M_3 are then merged into one single maximum $M_{1\_2\_3}$ as there is a valid path connecting these RIM.

To detect if valid paths exist for the merging condition, an image threshold is applied to each RIM at intensity, RIM – *d*. From the pixels above this threshold their corresponding *labels* are copied to a new image. Within this new image a neighbouring analysis detects connectivities between the RIM and its neighbour labels. This merging algorithm turned out to be time consuming as there are needed as many image thresholds as RIM.

Another solution to detect the valid paths for the merging condition is based on a *connectivity graph* analysis. Here a *graph* is created to represent the labels and the links among them, being modelled as *nodes* and *links*, respectively. The neighbours *links* are characterized by the maximum intensity found within their common boundaries (figure 3.37.(b)(c)), a value that will be used to evaluate the merging condition.

The *connectivity graph* is a list of *node* objects which contains for each: the *label*, the merging minimum value (*d*) and a list of object references and *link* intensities to its neighbour

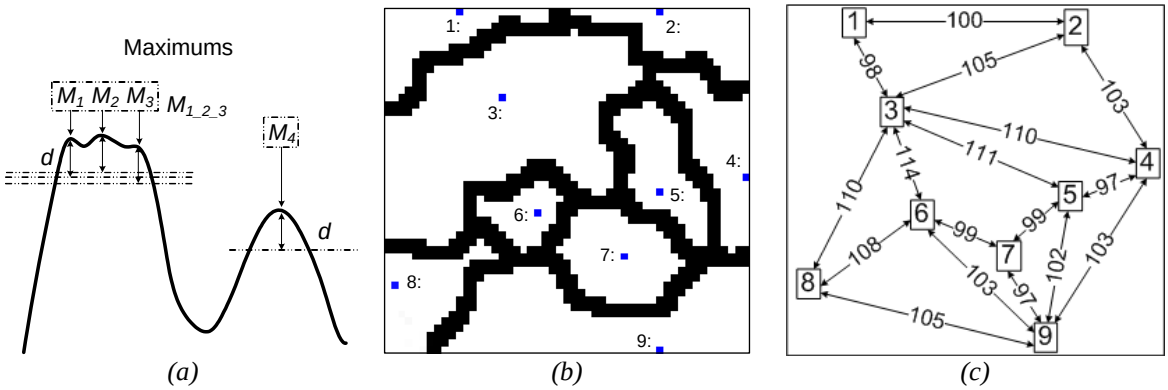


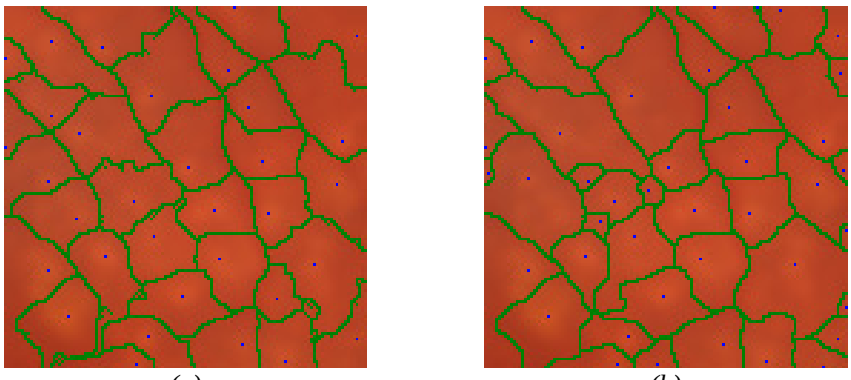
Figure 3.37 – GPL merging method. (a) two dimensional example of the merging condition; (b) *labels* image (blue dots are the RIM and black pixels are the region boundaries); (c) *connectivity graph* for the image in (b).

nodes. The merging operation for a *node A* will be to search in its neighbours' list the *links* verifying the merging condition and to merge with those verifying it. To optimize the search for mergeable neighbours an ordered search was implemented. The neighbours list is sorted by link intensity and is analyzed in a descending order. Once it is detected a link that is below the merging minimum value, the search is finished and the analysis proceeds to the next node.

To merge *node A* and *B* it is necessary to maintain the *A label*, add an entry converting *B* to *A* to the *equivalences' table*, update the maximum and minimum intensities and merge their neighbours' lists. Despite these main operations, to merge the neighbours' lists it is also needed to remove in *A* the link with *B*; add to *A* the new neighbour links from *B*; change in *B*'s neighbours the reference from *B* to *A*; and for the duplicated links: remove in the neighbour the reference to *B*, and update the value (*d*) with the new amplitude.

The merging operation using the connectivity graph analysis evaluates the *nodes'* list sequentially and when two *nodes* are merged the search will rollback to the first *node* in order to revaluate the new *connectivity graph*. This step finishes when the list is completely analysed.

In figure 3.38 it is shown the result of applying the GPL with the merging operation for an image with and without the *higher complete* transformation. By comparison with the previous solution without merging it is well noticed that the over-segmentation was reduced. Although when the algorithms are applied to the original image it appears a rough effect on the *plateaus* (some irregular borders on figure (*a*)). This could be improved using the *higher complete* image as it is shown on the figure (*b*), although the configuration of the value (*d*) is more complex as the amplitude scale will change from image to image depending on the number of plateaus. For the purpose of this GPL algorithm, which is to detect RIM, the solution without applying the *higher complete* transformation is the most adequate.



(a) *(b)*

Figure 3.38 – Image segmented using the gradient path labelling with:
(*a*) *nodes* merging (33 RIM) (*b*) *Image higher complete* and *nodes* merging (42 RIM).

Analysing the algorithm performance it was observed that the adoption of the *connectivity graph* in the merging operation resulted on significant decrease of execution time. For a 1M pixel image it was reduced by a factor of 30x, taking this algorithm less than 1 second to be completed (CPU: *Intel Core 2-1,86Ghz*; RAM: *2Gb*).

After applying the GPL and the merging operation the image spots are identified and characterized by their centre coordinates and amplitude. This will be the information used as the initial parameters for the modelling algorithm.

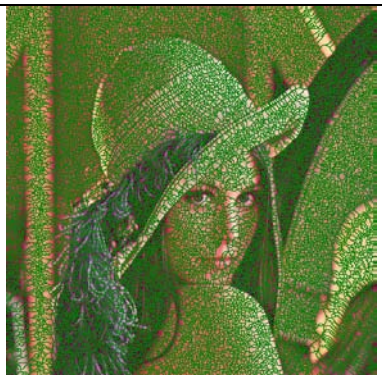
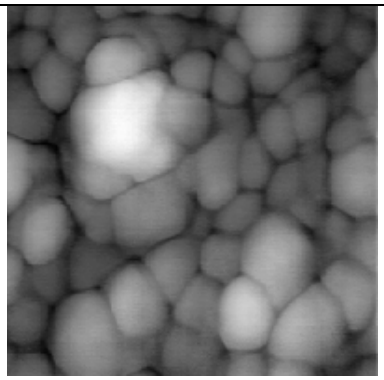
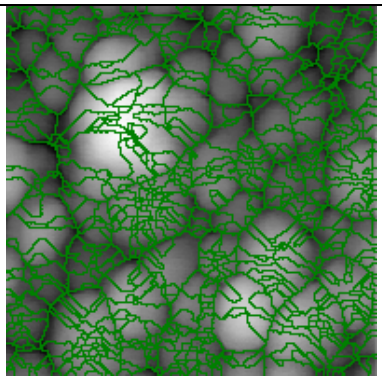
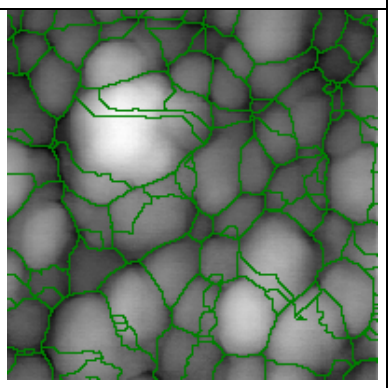
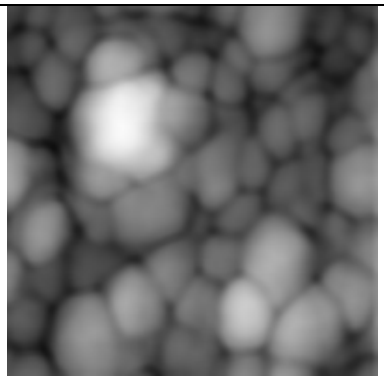
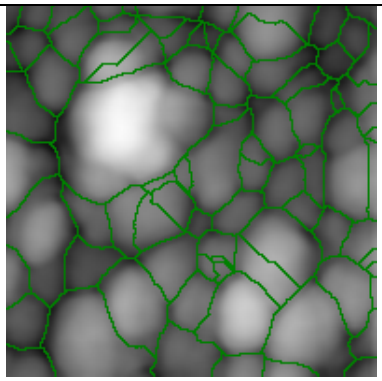
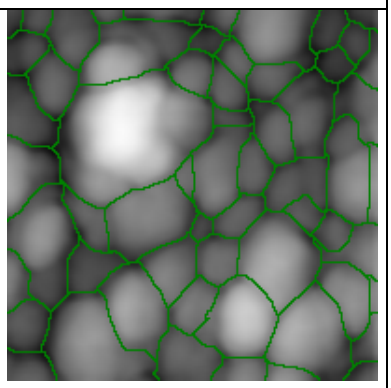
3.3.4. Considerations

The GPL algorithm presented is a segmentation algorithm adequate for the analysis of images containing an even or uneven illumination patterns. The gradient analysis provides an independence from illumination by searching the intensity changes and not just its amplitude. Also, the use of the image gradient to compute the image segmentation allows a reduction on noise artefacts, since the gradient is obtained by a pixel's neighbouring analysis.

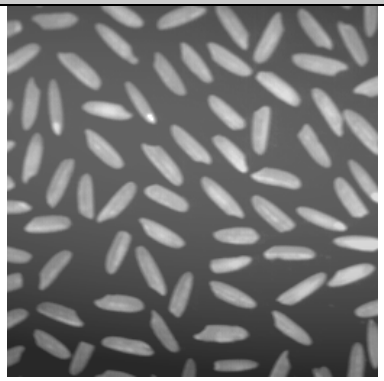
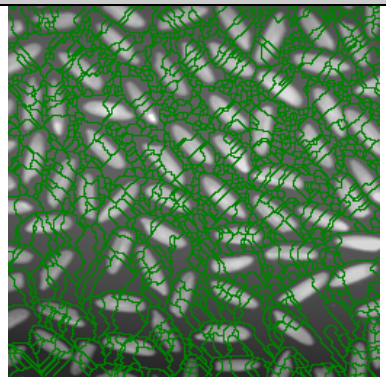
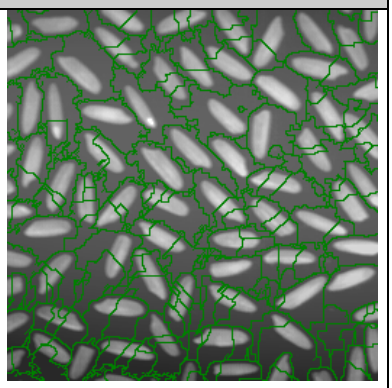
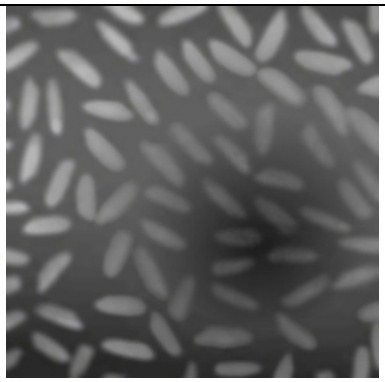
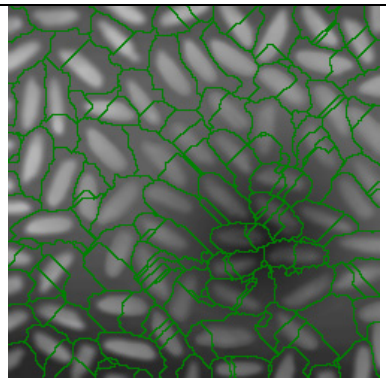
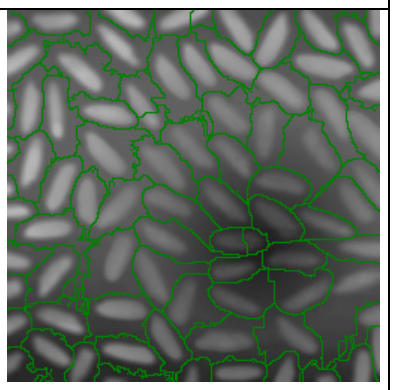
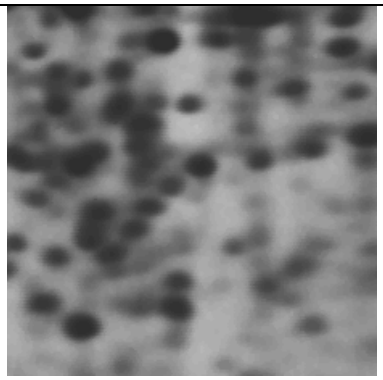
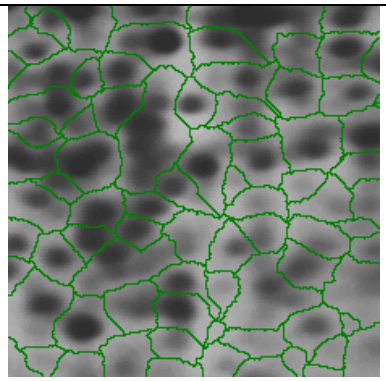
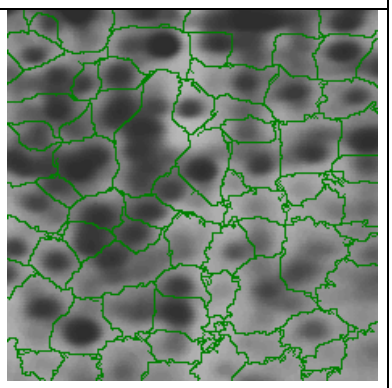
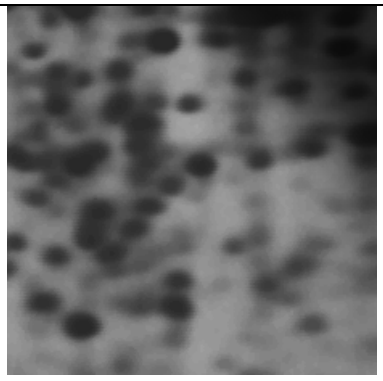
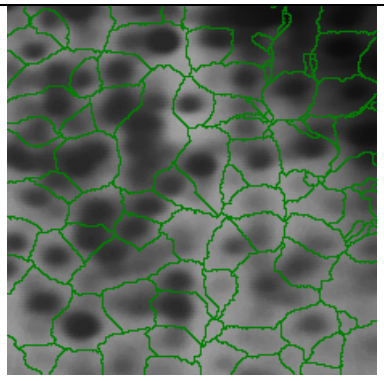
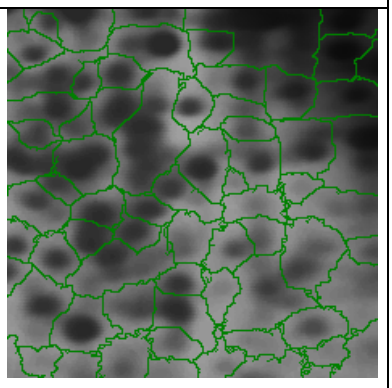
The Watershed Transform, which is the GPL most similar method, evaluates the intensity topology for segmenting the image. It has better independence from illumination in comparison to the thresholding methods, however it is also a method based on intensity amplitude where illumination deficiencies compromise the algorithm performance. When it is applied to the image without pre-processing for smoothing purposes it produces a significant over-segmentation.

A visual comparison between the GPL and the Watershed Transform is shown in table 3.1. The Watershed Transform implemented was the Hill Climbing algorithm without the *lower-complete* transformation. Both algorithms were applied to these test images without any image pre-processing, except the negative operation when required to detect dark or white objects. Typical image processing benchmark images (*cameraman*, *lena*, *rice*, *afmsurf* and others) were used to show the algorithms behaviour on these and to show their versatility. Some of them (*rice*, *Gel image*) were modified to contain uneven illumination and others (*afmsurf*, *rice*, *GEL image*) were smoothed in order to reduce the over-segmentation observed. The methods were also applied to retinal images containing different drusen sizes and densities. These latter images were tested with and without smoothing before applying the segmentation algorithms.

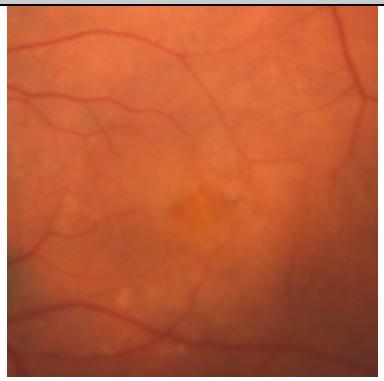
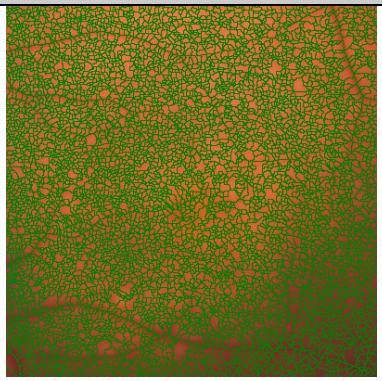
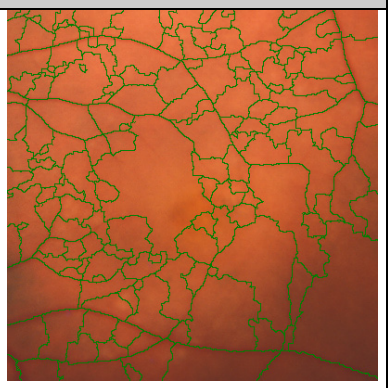
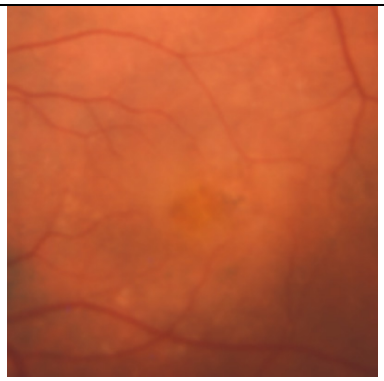
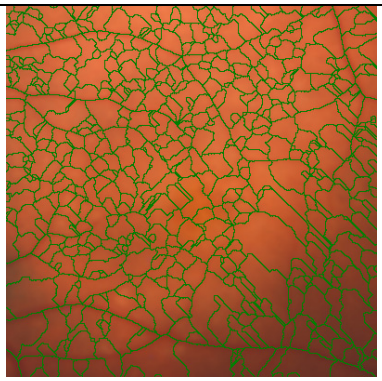
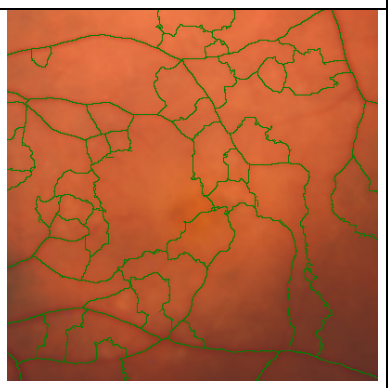
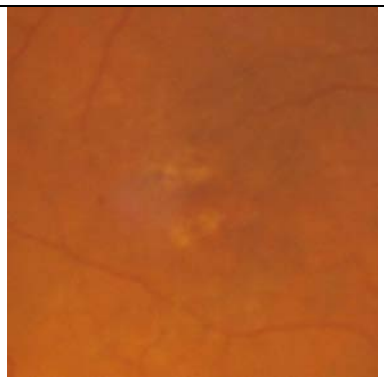
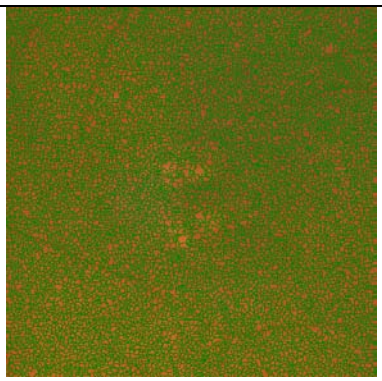
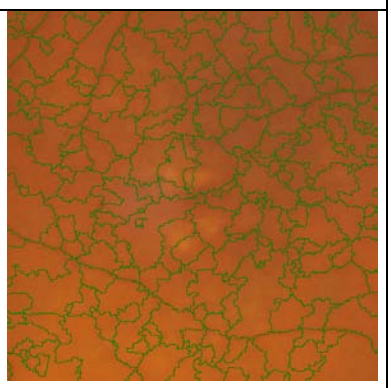
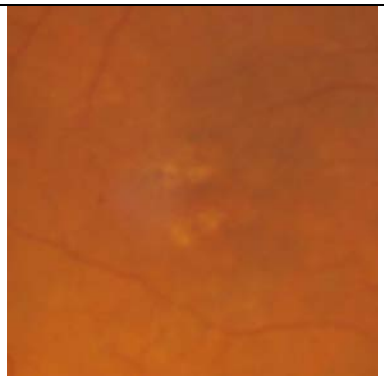
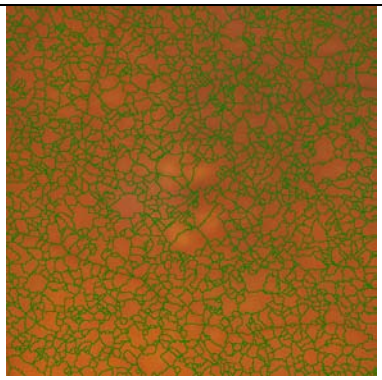
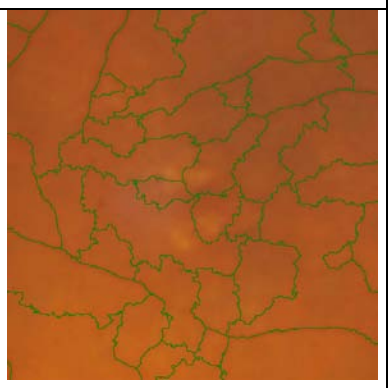
Table 3.1 – The Watershed Transform compared to the GPL algorithm.

	Original	Watershed	GPL
Cameraman			
Lena			
Afmsurf			
Afmsurf w/ 5x5 smoothing			

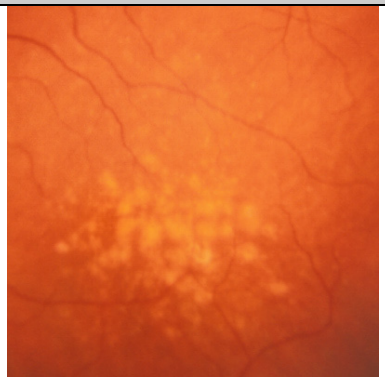
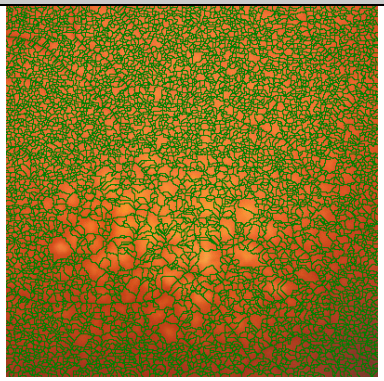
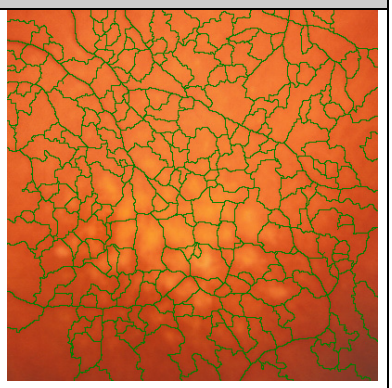
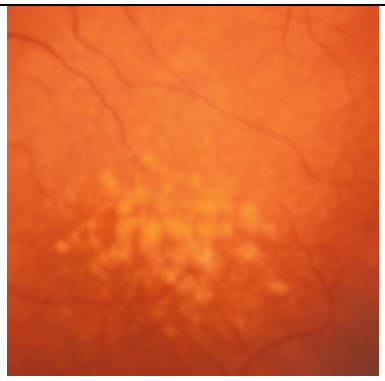
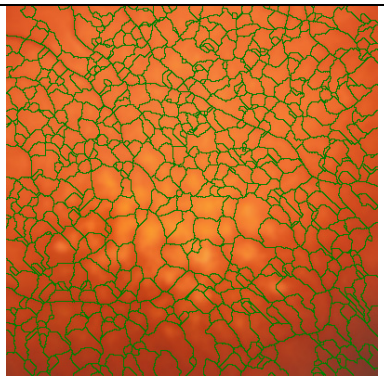
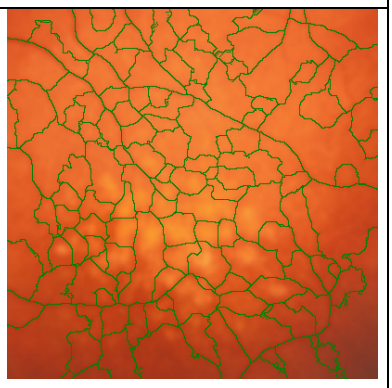
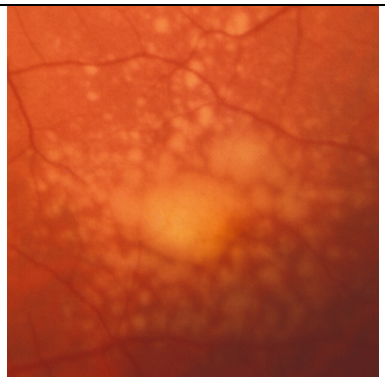
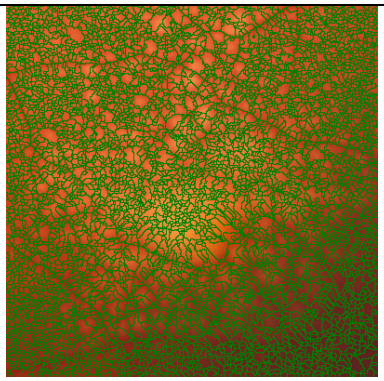
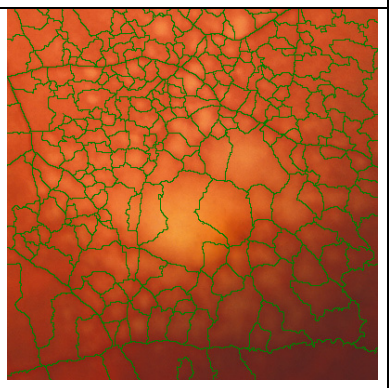
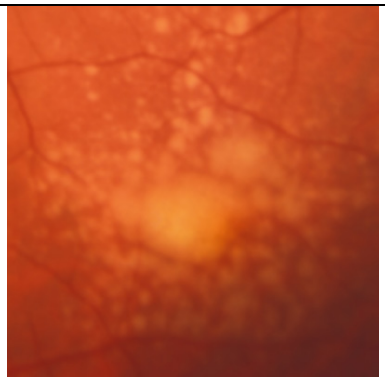
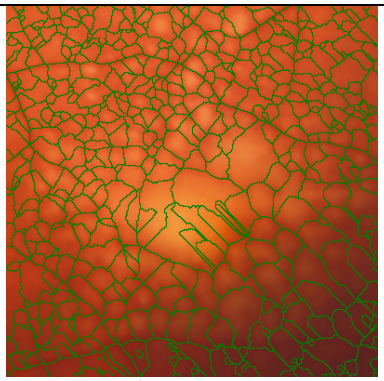
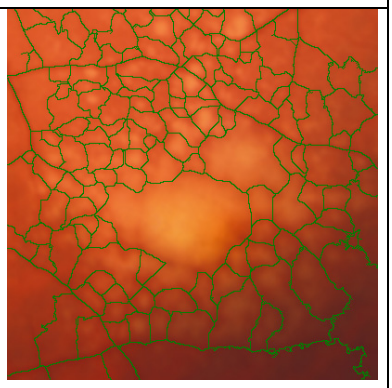
(cont.)

	Original	Watershed	GPL
Rice			
Rice w/uneven illum.+ smoothing			
GEL image w/smoothing			
GEL image w/ uneven illum.			

(cont.)

	Original	Watershed	GPL
Retina w/ small drusen			
idem + Smoothing			
Retina w/ small drusen			
idem + Smoothing			

(cont.)

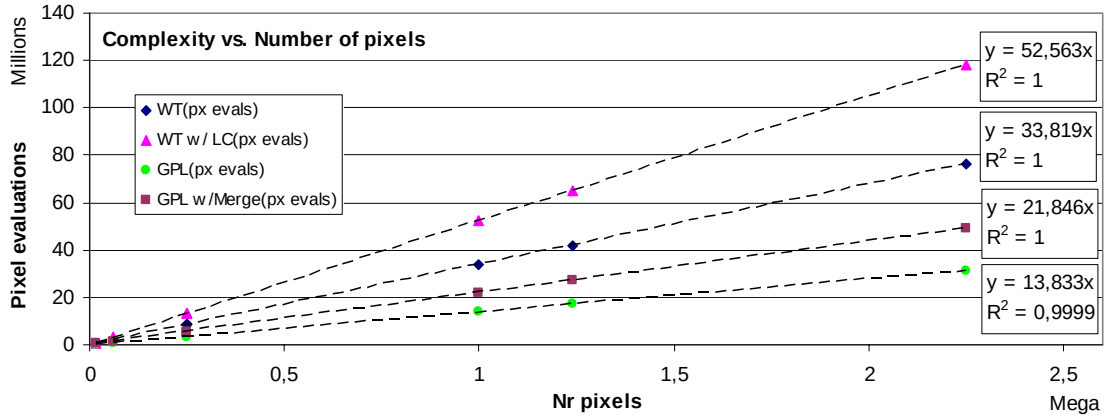
	Original	Watershed	GPL
Retina w/ medium drusen			
idem + Smoothing			
Retina w/ large drusen			
idem + Smoothing			

The algorithms have similar performances on less complex and smoothed images. Nevertheless, the Watershed Transform has higher tendency to over-segment the image. It should also be noticed the ability of both algorithms to detect spots even with non uniform illumination and lower contrast regions.

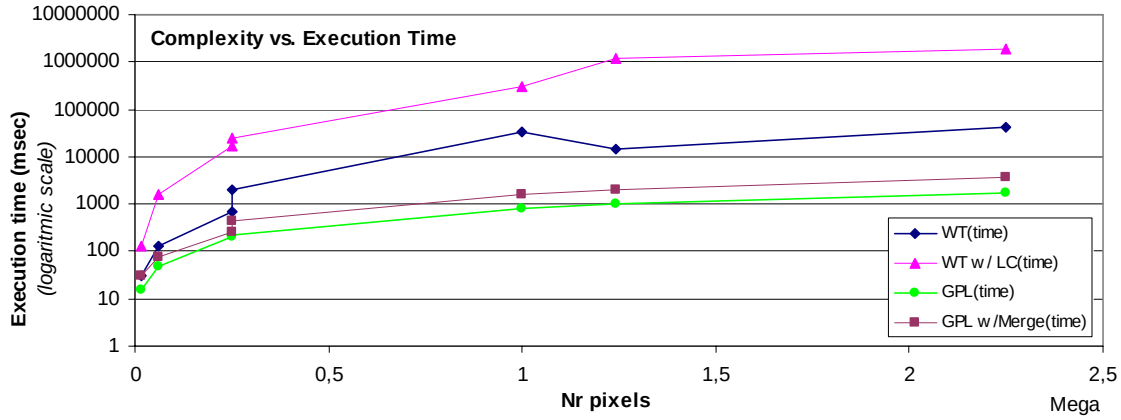
When comparing the algorithms' complexity there is the evidence that GPL has a faster execution, requiring less time to complete the detection than the Watershed Transform. Both algorithms have a complexity which is linear to the number of pixels, as it can be seen in figure 3.39.a. This figure shows the number of times a pixel is analysed in the Watershed Transform with and without the *lower-complete* operation and in the GPL algorithm with and without the *merge* operation.

The Watershed Transform requires ordered lists that do not affect its linear performance, since the insertion in the list is indexed and does not require sorting operations. However, the lists' manipulation required significant processing overhead, which has increased significantly the execution time, as it can be seen in figure 3.39.b. The *merge* operation used in GPL added extra complexity to the algorithm. Although its complexity depends on the number of objects found initially, the optimization introduced, which searches neighbour objects using a graph analysis, improved significantly the *merge* operation. The *merge* operation improves significantly the GPL algorithm results and only with the cost of doubling the execution time, which is not significant compared to the increase provoked by the *lower-complete* operation.

The features exhibited by the GPL algorithm, which provides robust object detection, showed that it can be applied to other image processing scenarios and not just retinal image processing. This is visible in the benchmark images shown in table 3.1 and it has been already tested for sunspots detection in solar images (Moreira 2008).



(a)



(b)

		WT			WT w/ LC			GPL			GPL w/Merge		
#	Nr Pixels	px evals	time	spots	px evals	time	spots	px evals	time	objs	px evals	time	spots
1	16129	545817	31	82	848456	125	82	235231	16	1022	365285	31	27
2	62250	2114044	125	143	3293463	1547	143	887125	47	597	1385722	78	83
3	249000	8457420	687	143	13175062	16906	143	3506412	204	884	5499296	265	84
4	250000	8431606	2047	4609	13089710	24032	4609	3569315	219	6986	5576301	454	436
5	1000000	33756942	32516	4609	52389334	303359	4609	13874124	828	15577	21889701	1594	453
6	1240996	41965728	14578	10269	65232021	1220422	10269	17345858	985	18282	27292108	1953	697
7	2250000	76121717	43422	8554	118340167	1908843	8554	30985930	1750	26883	49012813	3594	625

(c)

Figure 3.39 – Performance comparison between Watershed Transform and GPL (a) Graphical representation of WT and GPL's number of analysis of each pixel; (b) WT and GPL's execution time. (c) data details. (CPU: *Intel Core 2-1,86Ghz*; RAM: *2Gb*)

3.4 - Drusen Modelling

The next operation on drusen quantification is the identification of their boundaries in order to evaluate the affected areas. Especially in the soft type drusen, the boundaries are typically smooth with no defined contours. This characteristic makes the boundary detection a difficult task. The use of edge detection filters is not suitable as it does not only require good boundary contrast but, also contextual analysis to differentiate drusen from other structures such as vessels or macular holes (figure 3.40).

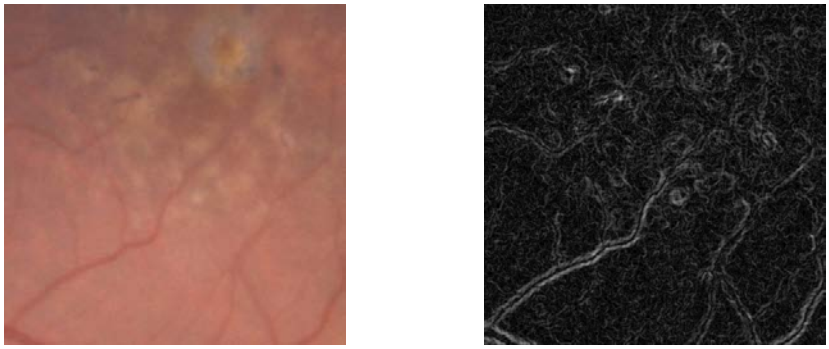


Figure 3.40 – Result of an edge detection filter applied to a retina region containing drusen and vessels.

Another approach that uses active contours (*snakes)* (Chan and Vese 2001) was also evaluated. Due to the parameterization complexity and the consequent difficulty to achieve convergent contours, this method was also considered not suitable for drusen detection. The same happened with the Discrete Wavelet Transform, which didn't show any relevant coefficients (figure 3.41).

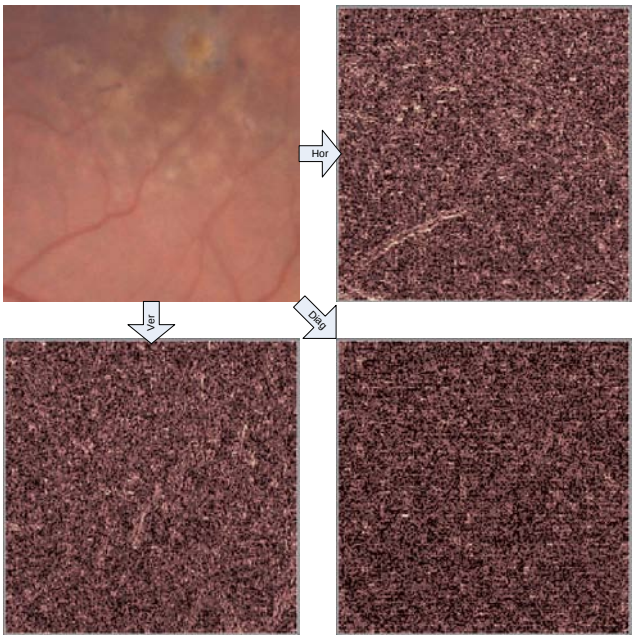


Figure 3.41 –Wavelet Transform applied to a retina image (Horizontal, Vertical and Diagonal Coefficients).

The adopted approach, involved the creation of an analytical model of the drusen found in the previous step. The model allows the shape and intensity characterization of drusen, providing information to evaluate them quantitatively. This information can be valuable to an alongside comparison that evaluates sequence of images during a long term treatment.

The small bubbles created by the accumulation of extra-cellular materials beneath the retina surface, appear on retinal imaging with approximately round or elliptical shapes and yellow colour. Also, viewing the fundus image as a topographical image, where the pixels' intensity represents elevations, drusen appear as topographical elevations. These observations motivated the creation of a model of the image intensity and from it extract drusen dimensions. The model is obtained by fitting analytical functions to the retina image.

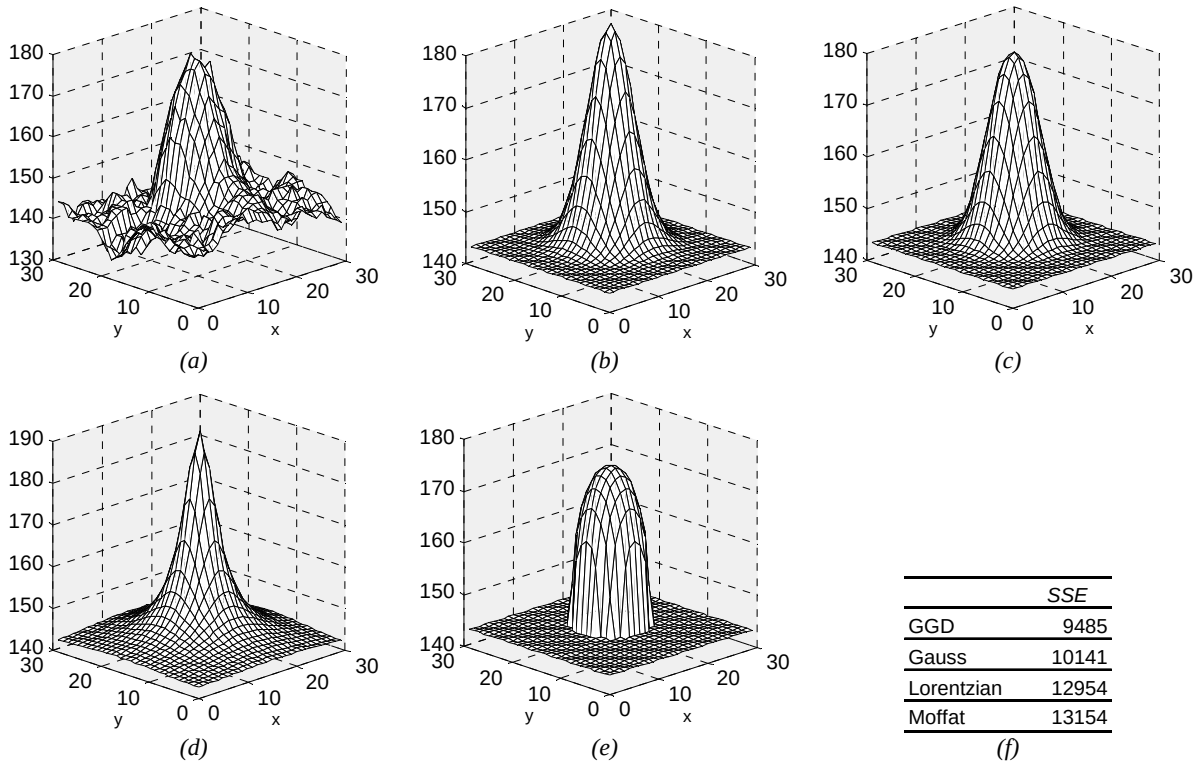
This approach can be a significant advance in Drusen detection algorithms since it provides not only a shape consistent segmentation and better noise independence when compared to threshold techniques, but also a good spot's characterization and reproducibility.

This technique is also used in other fields when the shape of the objects under detection is known. In spectroscopy it is used to evaluate the contributions from different spectral lines by fitting multiple analytical functions to the signal (Mora 2002; De Sousa Meneses, Gruener et al. 2005). It is also used in template matching for face detection (Attene, Falcidieno et al. 2006; Blanz, Scherbaum et al. 2007). In robotics it is used to create three-dimensional models of the scenes (Baeg, Hashimoto et al. 1995; Lester 1998; Biegelbauer and Vincze 2007). In medical imaging it is being used for surface rendering and for object segmentation (Song, Haralick et al. 2002; Jackowski, Satter et al. 2003). Also, in ophthalmology it is used in ocular MRIs to study the circularity of the retina (Atchison, Pritchard et al. 2005).

To adjust a model to an existing dataset it is necessary to choose how the model is described, i.e., what mathematical functions are used in the model, and also which method to be used to adjust the model parameters so that it fits the data. These two main steps are described in the following sections.

3.4.1. Function Model

The choice of the functions that constitute the model depends on their capability to describe the existing data. There should be a prior knowledge on the data distribution and eventually on their genesis to know theoretically what would be the correct shape. Also, if the data is coming from multiple sources, the model could be designed so that it can isolate the



	SSE
GGD	9485
Gauss	10141
Lorentzian	12954
Moffat	13154

Figure 3.42 – Drusen modelling (three dimensional views) (a) original image; (b) Gaussian function; (c) Generalized Gaussian Distribution function; (d) Lorentzian function; (e) Moffat function; (f) Sum square error between each function and the original image.

sources and therefore enable the study of individual contributions. In this study it is intended to find drusen contours and count the number of drusen if possible.

In a three-dimensional view of the retina surface as in figure 3.42.a, it can be seen a druse that has a waveform similar to a Gaussian function (figure 3.42.b) superimposed to a flat background with low amplitude noise. These observations and the fact that drusen have mostly rounded or elliptical shapes, has motivated the modelling of drusen by means of elementary functions.

This approach allows the drusen isolation from the background and their analytical characterization. Another advantage is that it removes noise from the image, while capturing only the drusen natural shape. Contributions from superimposed drusen can also be modelled together, generating a more precise model.

To choose the function which could model drusen, tests were carried out with the following analytical functions: the Gaussian, the Generalized Gaussian Distribution (GGD), the Lorentzian, and the Moffat (figure 3.42 (b,c,d,e)). These functions were fitted to the original image and the sum square error between them was evaluated. The analytical function which showed to model the drusen shape best, i.e., obtaining the lowest sum square error after fitting (figure 3.42.f), was the GGD function. This function, represented in equation (3.18),

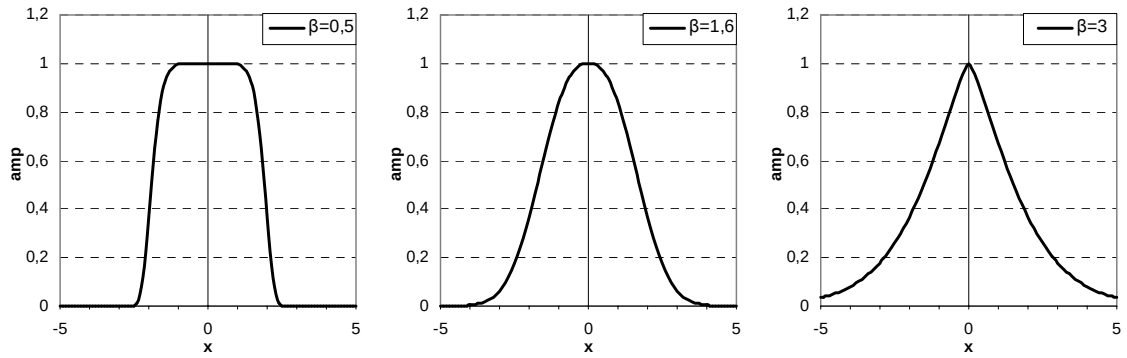


Figure 3.43 – GGD function profile in reference to three -β- values (A = 1; $x_0 = 0$; $s_x = 2$; $\theta = 0$).

uses 8 degrees of freedom, allowing shifts, rotations, width adjustments and profile modifications.

$$G(x,y) = A \cdot e^{-0.5*\left(\left(\frac{X}{s_x}\right)^2 + \left(\frac{Y}{s_x \cdot s_f}\right)^2\right)^{\frac{2}{\beta}}} + z_0 \tag{3.18}$$

where

$$\begin{aligned} X &= (x - x_0)\cdot\cos(\theta) + (y - y_0)\cdot\sin(\theta) \\ Y &= -(x - x_0)\cdot\sin(\theta) + (y - y_0)\cdot\cos(\theta) \end{aligned} \tag{3.19}$$

Amplitude scaling is obtained with the parameter -*A*-. Shift modifications are achieved in the *xy*-plane modifying the -x_0- and -y_0- parameters, while the parameter -z_0- shifts in the *z*-plane. The rotation is achieved by modifying the *x* and *y* coordinates by the angle -θ-. Width adjustments are allowed in *x* and *y* plane, by changing the width parameters -s_x- and -s_f-, respectively. The -s_f- parameter acts as a limiting shape factor between the width in x and y planes. These are the main parameters for a normal 3D Gaussian function.

In this work it was used a modified version which allows changes in the function amplitude profile. The profile parameter -β- controls the amplitude profile between square shape, bell shape and thin shape. Typical values for -β- vary between 0.5 to 1.6, knowing that such limits exhibit a profile as shown in figure 3.43.

The function partial derivatives with respect to all adjustable parameters will be needed for the modelling step. These were computed in Mathcad (2007) and are presented in the Appendix A.

As drusen are often side-by-side or even superimposed, the overall drusen model is composed by several individual GGD functions which are summed (3.20). The offset in z-

plane (z_0) is added at the end. The total number of seven parameters plus one offset must be determined for each druse.

$$Model\,(x,y) = z_0 + \sum_{drusens} A \cdot e^{-0.5*\left(\left(\frac{X}{s_x}\right)^2 + \left(\frac{Y}{s_x \cdot s_f}\right)^2\right)^{\frac{2}{\beta}}} \tag{3.20}$$

3.4.2. Modelling technique

To create an equivalent model of retinal drusen, the GGDs need to be adjusted to each of the previously detected drusen. This can be achieved by using non-linear least squares algorithms (Björck and Björck 1996) that adjust the function parameters so that it fits to the image intensity data, i.e., that it minimizes the sum square error (SSE) that is the sum of the squares of the differences between the model and the data in all *n* points (3.21). Several classes of algorithms are available for this purpose: Steepest Descent, Gauss-Newton, Genetic Algorithm's and others.

$$R(\mathrm{a}) = \sum_{i=0}^{n} [r_i(\mathrm{a})]^2 = \sum_{i=0}^{n} [y_i - f_i(\mathrm{a})]^2 \tag{3.21}$$

The Steepest Descent methods are iterative algorithms that use the function gradient to find the local minima. In each iteration new parameters (a_{new}) are updated using equation (3.22) in which the step value is proportional to the negative of the function's gradient ($\nabla R(a)$). The algorithm stops when the descent or the SSE is below a predefined tolerance. The algorithm advantage is that the search is always in the right direction as given by the gradient. However, its disadvantage is that the progression is slow when reaching the local minima, which could lead to a high number of iterations.

$$a_{new} = a - \gamma \cdot \nabla R(a) \tag{3.22}$$

The Gauss-Newton optimization method was derived from the Newton's method for finding roots of equations. The Newton's method can be applied in local minima and local maxima search since they are the derivative function's roots. The Newton's method is based on the analysis of the function's Taylor series. Considering that a change -Δ_a- in the parameter -*a*- can be approximated by the function linearization:

$$f_i(a + \Delta_a) \approx f_i(a) + f_i'(a) \cdot \Delta_a \tag{3.23}$$

where $f'(a)=0$ is the function's critical point for minimization, and the previous equation can be written as:

$$f_i'(a+\Delta_a) \approx f_i'(a) + f_i''(a)\Delta_a \tag{3.24}$$

the search direction given by -Δ_a- can be computed solving the equation

$$f_i''(a)\cdot\Delta_a = -f_i'(a)$$
$$\Delta_a = -\frac{f_i'(a)}{f_i''(a)} \tag{3.25}$$

which considers not only the first derivative, but also the second derivative providing a more convergent method.

$$f_i''(a)\cdot\Delta_a = -f_i'(a)$$
$$\Delta_a = -\frac{f_i'(a)}{f_i''(a)} \tag{3.25}$$

After solving the equation the next iteration on Newton's method is given by

$$a_{new} = a - \Delta_a \tag{3.26}$$

However the evaluation of the second-derivative can be a complex operation for some problems containing multiple parameters, since the Hessian matrix (3.27) containing the second-derivatives and cross-derivatives must be determined.

$$f_i''(\mathrm{a}) = H_i(\mathrm{a}) = \begin{bmatrix} \frac{d^2 f_i(\mathrm{a})}{da_1^2} & \frac{d^2 f_i(\mathrm{a})}{da_1 a_2} & \cdots & \frac{d^2 f_i(\mathrm{a})}{da_1 a_n} \\ \frac{d^2 f_i(\mathrm{a})}{da_2 a_1} & \frac{d^2 f_i(\mathrm{a})}{da_2^2} & \cdots & \frac{d^2 f_i(\mathrm{a})}{da_2 a_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2 f_i(\mathrm{a})}{da_n a_1} & \frac{d^2 f_i(\mathrm{a})}{da_n a_2} & \cdots & \frac{d^2 f_i(\mathrm{a})}{da_n^2} \end{bmatrix} \tag{3.27}$$

The Gauss-Newton method is from the class of methods that use approximations to Hessian matrix. The method is adapted to find the local minima from the SSE function, identical to equation (3.21). Here the critical point is where derivative of the residuals function is null.

$$\frac{dR(\mathrm{a})}{da_j} = 2\sum_{i=0}^{n} r_i(\mathrm{a})\cdot\left(\frac{dr_i(\mathrm{a})}{da_j}\right) = 0 \tag{3.28}$$

where

$$\frac{dr_i(\mathrm{a})}{da_j} = \frac{d\left(y_i - f_i(\mathrm{a})\right)}{da_j} = -\frac{df_i(\mathrm{a})}{da_j} = -\mathrm{J}_{f\,ij}(\mathrm{a}) \tag{3.29}$$

Evaluating the equation (3.23) and assuming that we want to adjust -*m*- adjustable parameters a_i; *i=1, 2 .. m*, the function derivative is its Jacobian:

$$f_i(\mathrm{a}+\Delta\mathrm{a}) \approx f_i(\mathrm{a}) + \sum_j \mathrm{J}_{f\,ij}(\mathrm{a}) \cdot \Delta a_j \tag{3.30}$$

The residuals are then obtained by:

$$r_i(\mathrm{a}+\Delta\mathrm{a}) \approx y_i - f_i(\mathrm{a}) - \sum_j \mathrm{J}_{f\,ij}(\mathrm{a}) \cdot \Delta a_j \tag{3.31}$$

To obtain the -Δa- the $r_i(a)$ of equation (3.28) must be replaced by equation (3.31) and the following -*n*- linear equations must be solved:

$$\begin{aligned} &-2\sum_{i=0}^{n}\left[y_i - f_i(\mathrm{a}) - \sum_j \mathrm{J}_{f\,ij}(\mathrm{a}) \cdot \Delta a_j \right] \cdot \left[-\mathrm{J}_{f\,ij}(\mathrm{a}) \right] = 0 \\ &\sum_{i=0}^{n}\sum_j \left[\mathrm{J}_{f\,ij}(\mathrm{a}) \cdot \mathrm{J}_{f\,ij}(\mathrm{a}) \cdot \Delta a_j \right] = -\sum_{i=0}^{n} \left[\mathrm{J}_{f\,ij}(\mathrm{a}) \cdot r_i(\mathrm{a}) \right] \quad \forall j = 1...m \end{aligned} \tag{3.32}$$

The previous equation when in the matrix format presents the following expression:

$$\mathrm{J}_f{}^{\mathrm{T}} \cdot \mathrm{J}_f \cdot \Delta a = -\mathrm{J}_f{}^{\mathrm{T}} \cdot R \tag{3.33}$$

This method is an improvement from Newton's method, since it is only dependent in function first derivative, and in the residuals which eases the finding of -*a*- parameters in the next iteration.

The other methods listed for function optimization are the Genetic Algorithms (GAs). These algorithms, initially proposed by De Jong (1975) and Holland (1975), are based on the same principles found on nature selection where a population's DNA is continuously evolving as the individual DNA combines with each other, in order to achieve better specimens. Here the problem's solutions represent the DNA, which are combined, suffering mutations and crossovers along the optimization process in order to find one good problem solution.

GAs are being used in several optimization problems where the goal is to find exact or approximate solutions using a pseudo-random search. Their main usage is to address optimization problems with many input parameters and possible solutions or exploratory problems where there is no complete knowledge of the problem being studied. In these cases the search space is very wide, which in other solutions presented will be probably non-convergent or long lasting to achieve an acceptable solution. Another advantage is that the search can be easily parallelized, reducing the time required to find a solution.

Some of their applications are on Bioinformatics (Zhang and Wong 1997; Ooi and Tan 2003), on exploratory scheduling applications (Wall 1996; Lim and Sim 2005), on electronic

circuits design (Kruiskamp and Leenaerts 1995; Zebulum, Pacheco et al. 2001) and on curve and surface fitting (Gulsen, Smith et al. 1995; Karr, Weck et al. 1995; Galvez, Iglesias et al. 2007).

In *GAs* the search for a set of parameters that solve a specific problem is an iterative process that makes pseudo random changes in the parameters, in order to maximize its' fitness function. The algorithm initially creates several sets of start up parameters which are candidate solutions, i.e., the start-up population of DNA chromosomes, and evaluate their fitness. The population evolvement in the direction to a better fitness is made by *selecting* the candidate solutions which obtained a good fitness, *combining* them (*crossover mechanism*), generating new solutions, and randomly *mutating* some of the candidates to generate diversity in the population. At the end of each iteration, it is applied the culling mechanism to maintain the size of the population and keep those solutions that have more potential to improve the fitness function. The search ends when the desired termination criteria is met, which can be a fitness threshold or a number of iterations.

In functions optimization or surface fitting, which is the case of the drusen modelling, the initial population are sets of function parameters in their approximate location. Gulsen, Smith *et al.* (1995) presented a method to apply GAs on curve fitting, which could be generalized to surface fitting. Here the target is to minimize the fitness function, as presented by the residuals function in equation (3.21). The chromosomes are the parameters themselves and the *crossover mechanism* is achieved by the arithmetic mean of two solutions of parameters. The *mutation mechanism* here is a perturbation provoked by a change in the parameters values, which occurs within a range of possible pre-defined values. Since this is a perturbing operation, mainly used to generate new fronts of parameters, it is only applied to random iterations and to random chromosomes.

In drusen modelling the major difficulty is the function and residuals evaluation and not the parameters generation itself. The model's derivatives are known and consequently the search direction too, which has limited the search space. Also, the number of pixels is usually significantly higher than the number of parameters, making the function evaluation the most expensive operation in each iteration. From there on, in each GAs iteration, there are dozens of function evaluations. Depending on the size of the population, the GAs will be much more time consuming then the other presented methods. Therefore, the GAs were not considered suitable to surface fitting nor for drusen modelling.

The two first classes of methods presented (Steepest Descent and Gauss-Newton) had each of them their own advantages. While the first is good for the initial search and for the

alternative solutions search, the second is more adapted to the search refining. Levenberg (1944) introduced a "*damped version*" of the Steepest Descent and Gauss-Newton's method as shown in equation (3.34).

$$\left[\mathrm{J}_f{}^{\mathrm{T}} \cdot \mathrm{J}_f + \lambda \cdot \mathrm{I}\right] \cdot \Delta a = -\mathrm{J}_f{}^{\mathrm{T}} \cdot R \tag{3.34}$$

Later with the work of Marquardt (1963) the optimization method was further improved to be faster. He replaced the identity matrix by the diagonal of the Hessian approximation ($\mathrm{J}_f{}^{\mathrm{T}} \cdot \mathrm{J}_f$):

$$\left[\mathrm{J}_f{}^{\mathrm{T}} \cdot \mathrm{J}_f + \lambda \cdot diag\left(\mathrm{J}_f{}^{\mathrm{T}} \cdot \mathrm{J}_f\right)\right] \cdot \Delta a = -\mathrm{J}_f{}^{\mathrm{T}} \cdot R \tag{3.35}$$

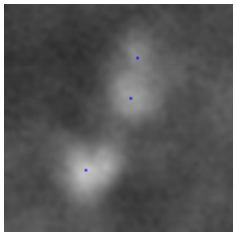
Depending on parameter -λ- the expression will perform closer to the first or the second method. This parameter will be adjusted during the iterative process depending on the evolution of the residuals.

During the iterative process, if a fast decrease on the residuals is observed the damping parameter can be decreased, bringing the algorithm closer to the Gauss-Newton method. In the opposite case, if the residuals were insufficiently decreased, the damping parameter can be increased, taking the algorithm closer to a Steepest Descent method. This last operation can be useful to prevent the algorithm from getting stuck in local minima.

There is no exact heuristic to find the correct step-up and step-down values for the damp parameter. The recommendations are an initial value for λ below 1 to start with a Gauss-Newton method when the solution is near and a value above 1 when the solution is not known or is far. For step-up λ is multiplied by a step factor *v* and for step-down λ is divided by the step factor *v*. The step factor *v* defines the sensitivity of the algorithm and usually assumes a factor of 10. A large value will generate higher changes at each iteration and may step over the function minima and diverge, while a low value can make the algorithm get stuck on local minima.

For drusen modelling this was the chosen method. It is capable of adjusting multiple function parameters and computing multiple function superimpositions while pursuing the best SSE between the model and the image, which were the requisites initially established.

The initial step in modelling stage is the *parameters initialization.* As it was previously mentioned the model to be fitted is composed by one GGD function for each detected drusen plus the offset -z_0- in the *z*-plane. At this stage, these functions are initialized with the location (x_0, y_0) and amplitude retrieved by the drusen detection algorithm presented on section 3.3.3, while the offset in *z*-plane -z_0- starts-up with the mean intensity value of the



	Parameters							
	#	A	X0	Y0	Sx	Sf	β	θ
Functions	1	57	59	24	5,5	1	0,8	0
	2	76	56	42	5,5	1	0,8	0
	3	111	36	74	5,5	1	0,8	0
							Z0	82,8

Figure 3.44 – Example of GGD functions parameters initialization.

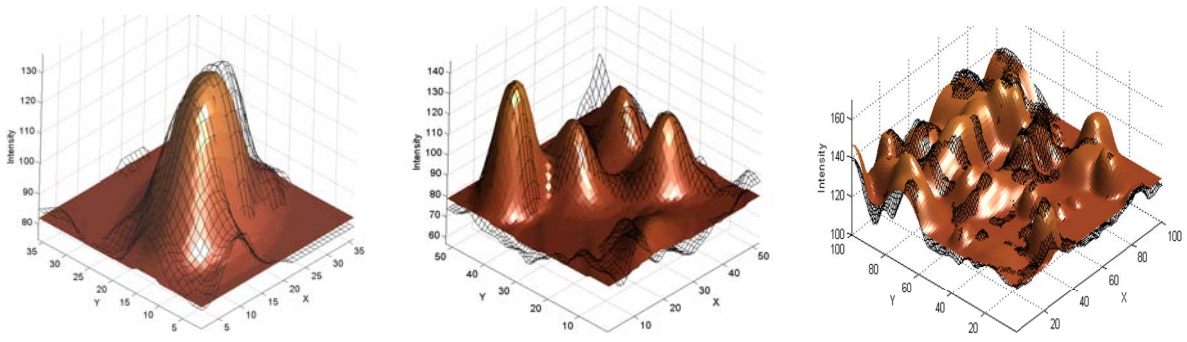
image. These GGD amplitudes are obtained by subtracting the -z_0- to the RIM intensity. Finally, the other function parameters (width, shape factors and rotation) are set to predefined values, obtained by averaging the results from a training test set. This initialization is fundamental to reduce the complexity of the fitting algorithm and to increase its probability of convergence. An example of initialization is presented in figure 3.44.

The second stage is to run the Levenberg-Marquardt fitting algorithm in order to adjust the model parameters so that it fits the image data. Since the model is initialized, but not near the optimal solution, the λ damp parameter was initially set to 1 which is a balance between both algorithms, while the step value *v* maintained the recommended value (10).

The stopping condition was that a maximum number of iterations is exceeded or the λ exceeds a maximum value. The latter is obtained by multiplying the step value *v* by a constant, which is the minimum number of consecutive iterations above λ=1 without decreasing SSE.

Figure 3.45 shows a three dimensional view of some examples of real images in comparison to their analytical model. This model allows a better separation of the noise from the drusen in analysis.

In figure 3.46 and figure 3.47 are shown the initial stage of the fitting progression for some examples. The small images only show the iterations where a successful reduction on the residuals is observed. The plots show the progression of the functions parameters along



(a) *(b)* *(c)*

Figure 3.45 – Examples of drusen modelling viewed in three-dimensions. (a) isolated druse; (b, c) superimposed drusen. (*grid* – original image; *surface* – analytical model)

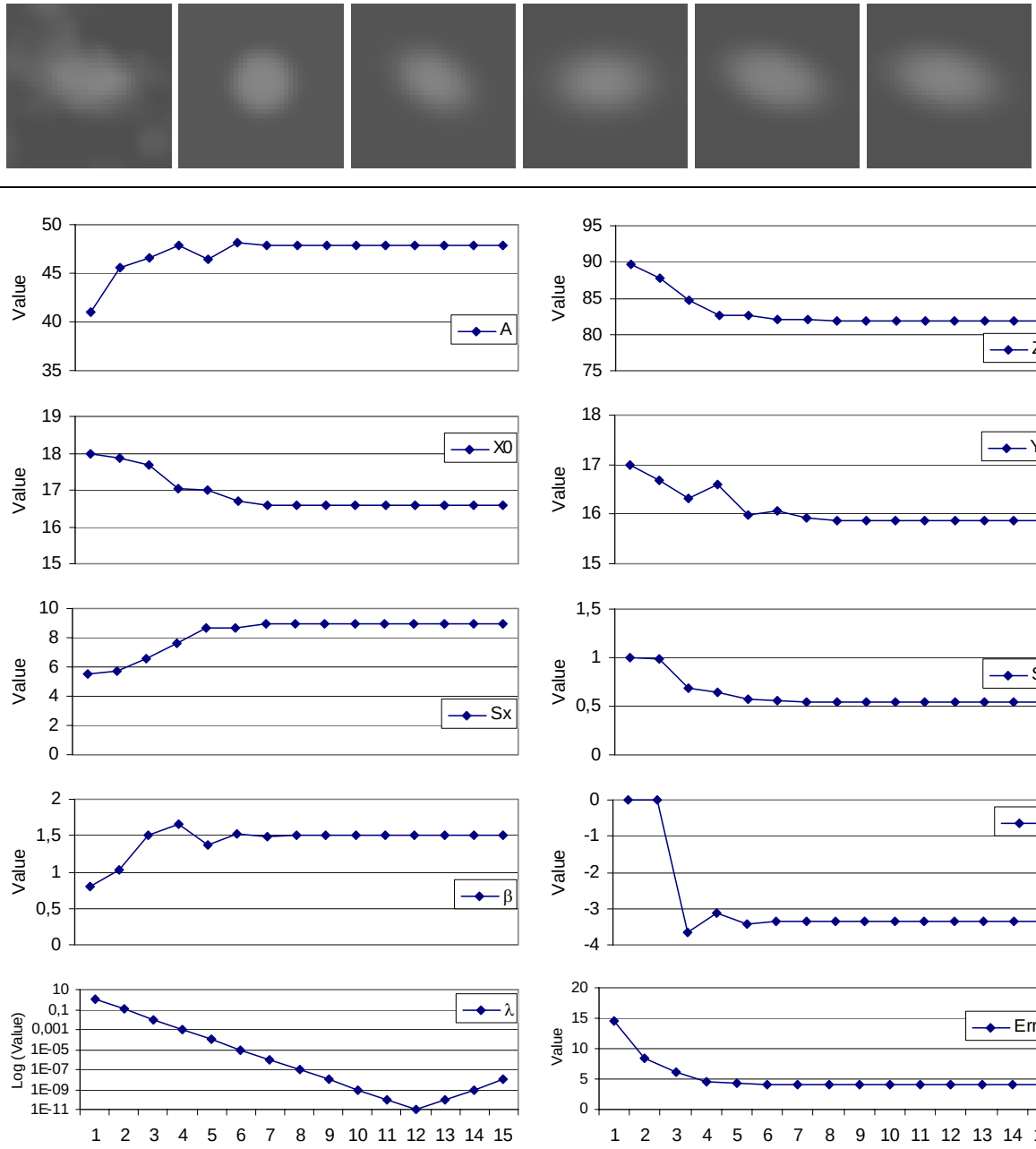


Figure 3.46 – Parameters progression for fitting image 1.

the fitting process. The two plots at the bottom present the damp parameter and the SSE. The latter was defined as the average SSE by pixel. It can be observed that after a small number of iterations the result is near the final solution, even on more complex images such as those shown in figure 3.47.

The algorithm's progression during an optimization process starts with an initial usage of the Gauss-Newton method and when the residuals get to a local minimum the damp parameter is increased, changing gradually to a Steepest Descent in order to find alternative solutions. When a solution that decreases the residuals is found, it returns progressively to a Gauss-Newton until the solution stabilizes. However, at the end, when a strong local

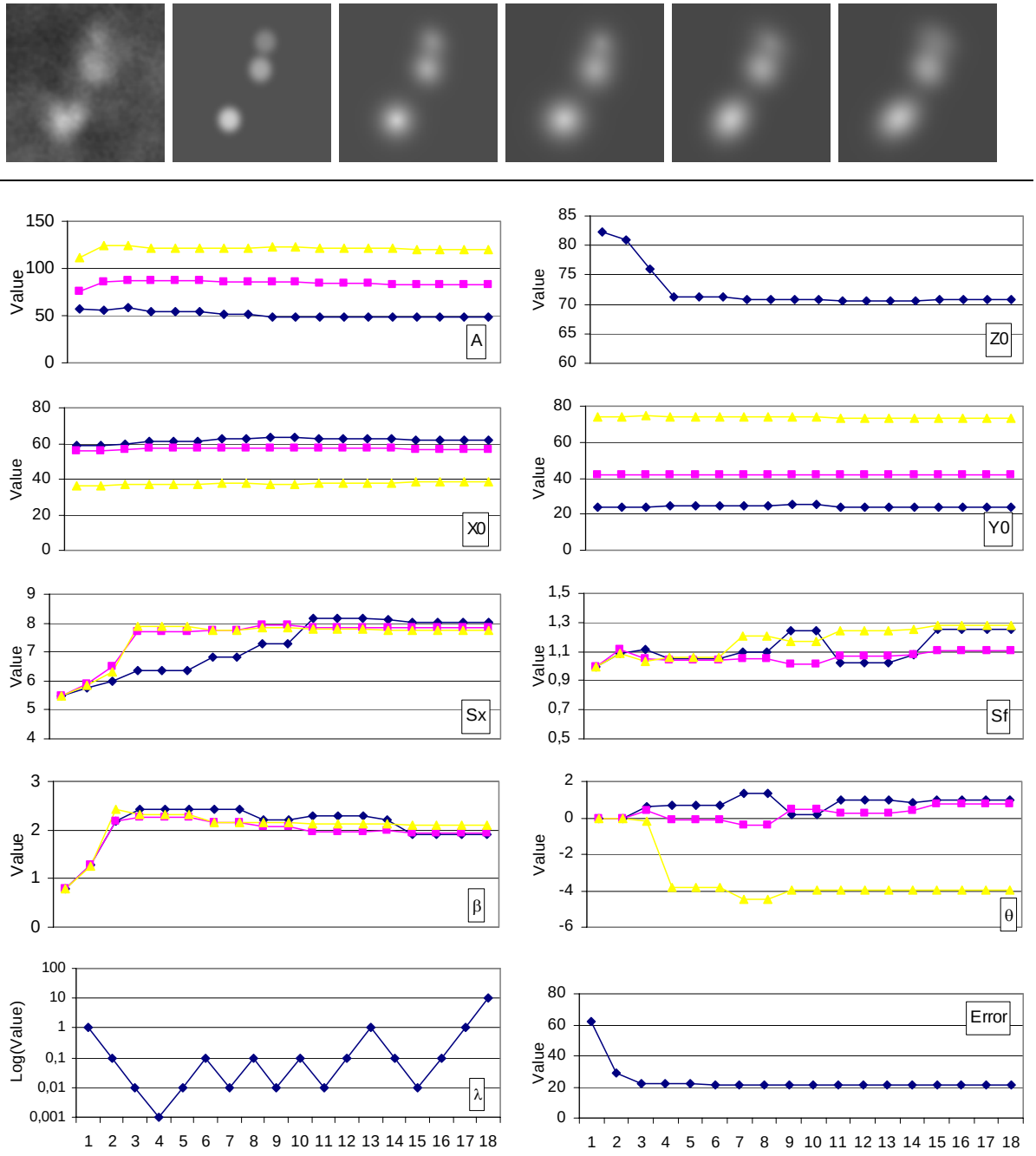


Figure 3.47 – Parameters progression for fitting image 2.

minimum is found the damp parameter will increase and finish as a Steepest Descent, while searching for alternative solutions.

In order to limit the search space and to create valid and coherent models it was introduced a constrained version of the Levenberg-Marquardt optimization algorithm. This operation is especially important for the shape factor, since it is expected to have GGD functions with approximately rounded shapes. These constraints can be defined with absolute or relative bounds. For location and amplitude parameters they have relative bounds, while for the others the generated values must be inside an interval of absolute values.

Whenever a parameter calculated by the Levenberg-Marquardt algorithm falls outside the constraints, it will be replaced by its nearest valid value, i.e., the upper or lower bound. However these boundary constraints are not needed to be set for all parameters, as it is influencing the algorithm progression and might difficult the convergence.

Considering the tests performed, the previous method was only recommended for the shape factor -s_f- and Gaussian profile -β- parameters. The first parameter was set with a range of [0.25 .. 4] and the second with a range of [0.3 .. 3].

The Levenberg-Marquardt algorithm can be summarized as follows:

1: Initialize all the GGD function parameters {A, x_0, y_0, s_x, s_f, β, θ}, the offset in the z-plane z_0 and the damp parameter λ.

2: Evaluate the SSE between the image and the initial model.

3: Solve the linear equations to find the new set of function parameters.

4: Evaluate the SSE between the image and the new model and:

i. divide λ by the step value *v* and maintain the new parameters, if the error decreases;

ii. multiply λ by the step value *v* and reject the new parameters, if the error increases.

5: Stop if λ exceeds the upper limit or if a maximum number of iterations is achieved, or go back to *step 3.*

3.4.3. Modelling Performance Optimization

The creation of the drusen model for the macular region can be a compute intensive operation. The macula can contain a large number of drusen and consequently the model will contain a large number of GGD functions, added to the fact that each of them contains seven adjustable parameters, which constitutes a large search space problem that is highly computed intensive to solve. Also, the complexity for solving the Levenberg-Marquardt equation (3.35) is quadratic in order of the number of pixels (*n*) and the number of parameters (*m*), resulting in a complexity of $O(n^2 \cdot m^2)$. For all (*x*,*y*) locations, GGD functions and partial derivatives must be computed followed by the convolution of both matrixes. This compromises the overall performance as these two indicators are usually related, i.e., on a wider image the probability of having more drusen increases.

These observations were tested when the whole macular region was modelled directly. The set of retinal images used in this study contained images with a size of 1 to 2 Mega pixels and some contained dozens of drusen. The result was a long optimization process that took

hundreds of iterations to find a stable solution. Note, also, that one iteration can last an average of 5 minutes to solve on a computer[1]. Therefore the consequence is that this process is not usable.

To reduce the complexity of the fitting process two modifications were introduced. The first was based on the fact that the image resolution is sufficient to be analysed in a sparse manner. The second consisted in *sectioning* the image into smaller images containing only isolated drusen which are fitted separately, reducing both the number of pixels and the number of GGD functions.

The first modification is to evaluate the GGD functions only on a reduced number of points. Because the image resolution allows it, the model can be extracted from a mesh of points sparse on space. The number of evaluations was decreased 50% or more since the pixels' spacing used to evaluate the model was increased to 2 pixels in the x-axis and the y-axis. To study the sensitivity of the fitting algorithm regarding the pixels' spacing, several images containing different number of drusen were evaluated. For each one the SSE, the *sum absolute error* (SAE) and the average execution time was analysed with a pixels' spacing between 1 (original) to 4 pixels. The sensitivity test results are presented in table 3.2. Here it can be concluded that, until a spacing of 3 pixels, the average execution time of the fitting algorithm decreases significantly without compromising the precision (SSE and SAE).

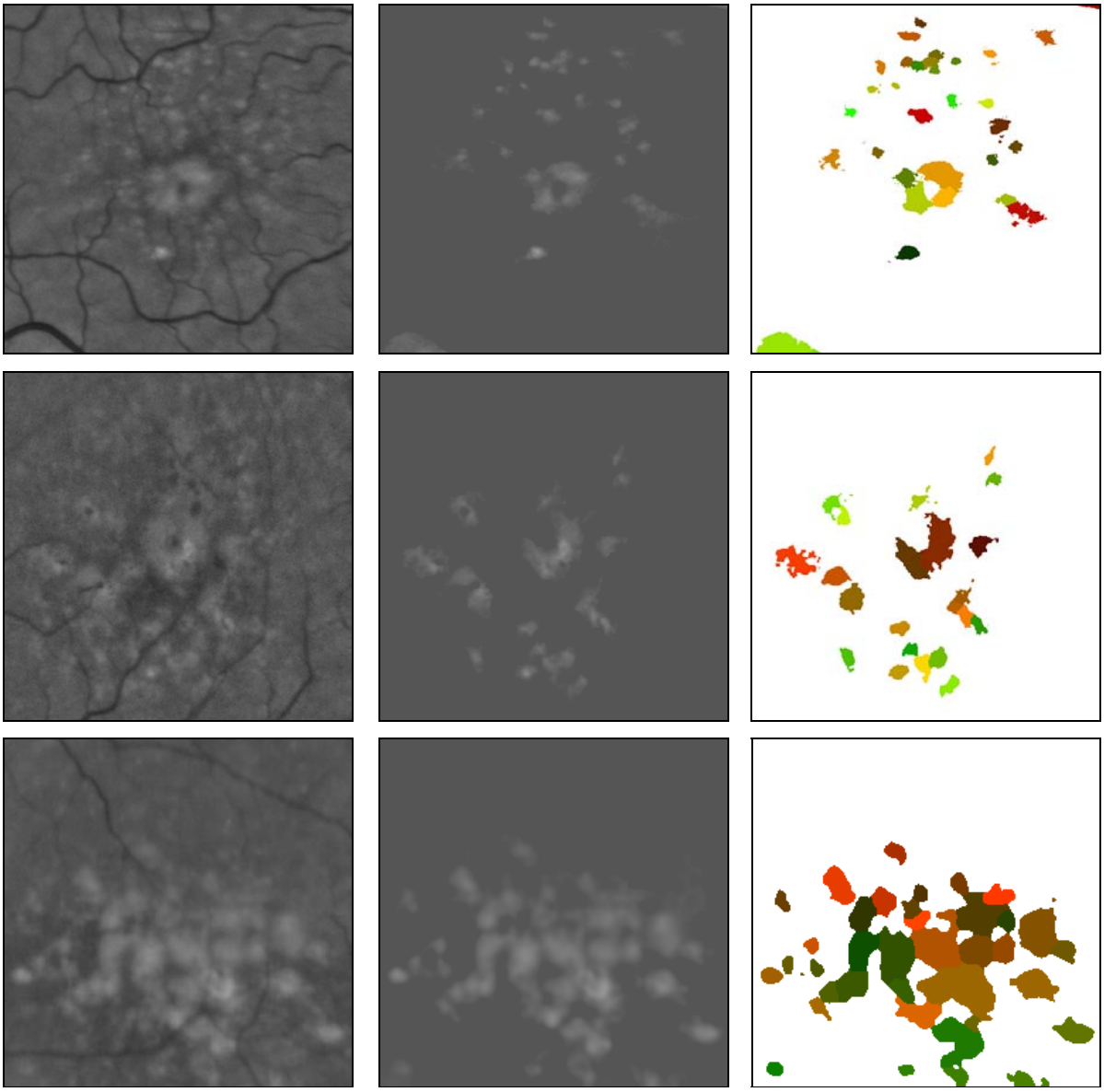
Table 3.2 – Performance of the fitting algorithm in terms of SSE, *SAE* and average execution time for different pixels' spacing.

Number of Drusen	Spacing (pixels)	SAE	SSE	Iters	average time per iteration (ms)
1	1	1,43	12,58	19	20,58
	2	1,05	8,26	40	11,33
	3	1,67	8,78	37	11,41
	4	2,26	17,81	44	10,30
4	1	2,01	13,99	21	154,00
	2	1,93	13,14	51	47,78
	3	1,96	13,43	19	32,89
	4	1,98	13,63	51	21,43
8	1	1,47	14,41	28	2089,29
	2	1,51	15,04	33	569,70
	3	1,52	15,07	18	263,89
	4	1,46	13,93	21	165,95
18	1	2,85	20,88	25	9332,52
	2	2,76	19,26	34	2971,97
	3	2,76	19,43	32	1309,06
	4	2,88	20,01	24	836,58
22	1	2,32	18,11	51	31405,94
	2	2,30	17,05	32	8437,50
	3	2,32	17,57	24	4041,67
	4	2,32	17,74	30	2608,87

[1] Tests on a CPU: Intel Core 2-1,86Ghz; RAM: 2Gb

The aim of the second modification was to create smaller images which contained drusen or superimposed drusen to be processed afterwards. This modification has also the added value of allowing the parallelization of the modelling step, which can decrease significantly the processing time. The *sectioning* purpose is to identify which isolated drusen can be modelled independently, and which superimposed drusen could be modelled together but, in a smaller image.

The *sectioning* is constructed with a combination of a threshold operation and an output from the Gradient Path Labelling presented in section 3.3.3. In order to identify whether the drusen are isolated or superimposed, first, an image is created containing the area



(a) *(b)* *(c)*

Figure 3.48 – Initial stage of the sectioning method for three examples. (a) original image; (b) image after GPL; (c) marks threshold.

of influence from the selected RIM displayed over a uniform background (*intensity* = 85), as shown in figure 3.48.b. Then a threshold is applied to this image with a value of 10% above the normalized background (*intensity* = 93). The pixels whose intensity value is above the limit are replaced by their *marks* (figure 3.48.c). This step will show which drusen are isolated, i.e., surrounded only by background, and those that don't, which have adjacent RIM. The objects surrounded by background (isolated or superimposed groups) are then copied to new images containing a predefined margin of background pixels to avoid contact with the image borders (figure 3.49). This strategy implies that only drusen above this threshold will be considered significant to be analysed, which is generally accepted by ophthalmologists.

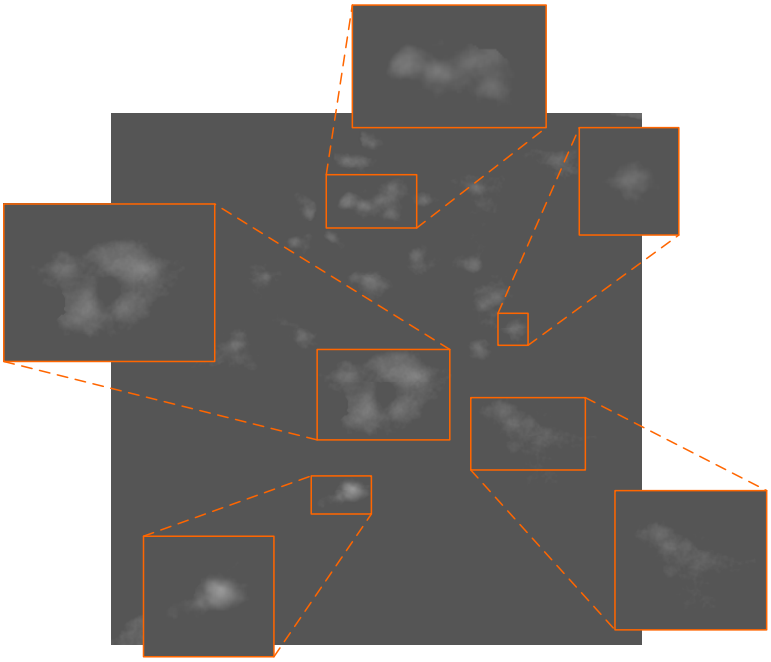
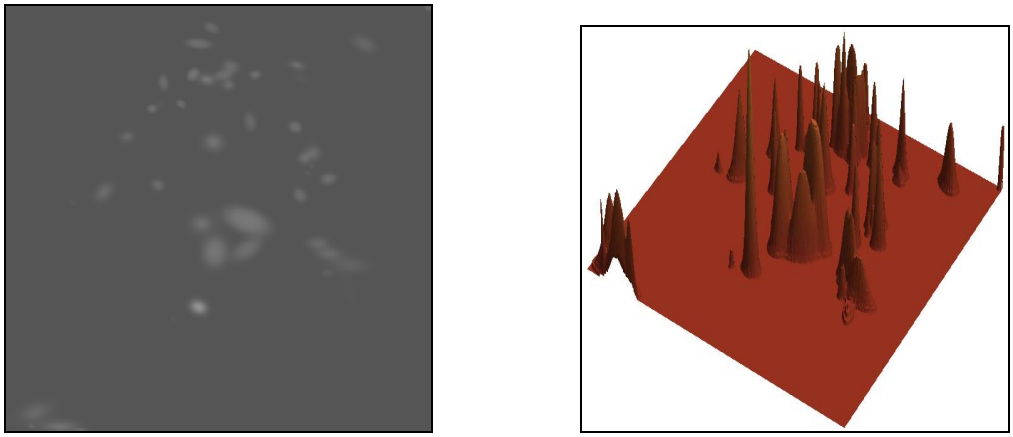


Figure 3.49 – Output from the image *sectioning* and from the modelling step.
The zoomed areas exemplify some of the sectioned images.

The sectioned images are then analyzed with the modelling algorithm, resulting in several sets of GGD functions that combined represent the image model, as shown in figure 3.50 (a,b). The image on the left is the natural two-dimensional representation, where the pixels' intensity represents the GGD function values. On the right, the three-dimensional view shows the GGD functions shape. The profile looks very thin and tall but this is just an optical illusion provoked by a narrow *z*-scale.

3.4.4. Considerations

The modelling step now presented, introduced a novel approach to the retinal image analysis. The principal basis is that the drusen have a typical shape that can be approximated



(a) (b)

Figure 3.50 – Output from the modelling step. (a) 2D view (b) 3D view.

by an analytical function. This approach allows a shape consistent segmentation which decreases noise and border irregularities influence.

The analytical function chosen (*Generalized Gaussian Distribution*) allowed its adjustment in eight degrees of freedom: amplitude (A), location (x_0, y_0), rotation (θ), widths (s_x, s_f), amplitude profile (β) and offset in z-plane (z_0). This function is a modified version of the Gaussian function to which was added the amplitude profile parameter, which allowed thinner or wider profiles. The result is a function that is well adapted to represent retinal drusen.

The Levenberg-Marquardt algorithm used to fit the model to the image is a well known and proved method for the optimization of non-linear functions. The algorithm is able to fit superimposed functions based on the analysis of the SSE between the model and the image. The algorithm is very robust and flexible, since it is a conjunction of two other well known optimization methods (Steepest Descent and Gauss-Newton) which can behave as one of two, depending on the optimization evolvement.

The major drawback from applying this method directly over the whole image was the intensive computation. The method has a quadratic complexity in order to the number of pixels and to the number of parameters to adjust. This can lead to a long lasting fitting operation and eventually not converge as the search space is very wide. The solution to this problem came with the introduction of two modifications. The first was to make the analysis in a sparse manner, reducing artificially the image resolution. The sensitivity tests showed that this reduction in the resolution didn't affect the fitting precision. The second modification was to divide the problem complexity into several smaller problems which are less complex to solve, since it reduces the number of pixels and the number of parameters as well. The method consisted in copying isolated or superimposed drusen to smaller images where the

modelling is then applied. The result is a set of small images containing a small number of drusen which is faster to fit.

3.5 - Drusen Quantification

With the retinal drusen model it is now possible to proceed to the drusen quantification. In the epidemiological studies which have been done in the past, the calculated indicators were mainly the number of drusen and the size of the biggest druse present, but only within a range of four to five possible values. However, a high variability among the analysis done by different specialists was observed. In the Wisconsin Grading Center, considered as one of the reference Retina Image Grading Centres, a variability of *32.9%* for drusen size measurement was reported (Klein, Davis et al. 1991).

The quantification method that is proposed here quantifies the affected area, estimates the drusen integral and is able to evaluate the number of drusen.

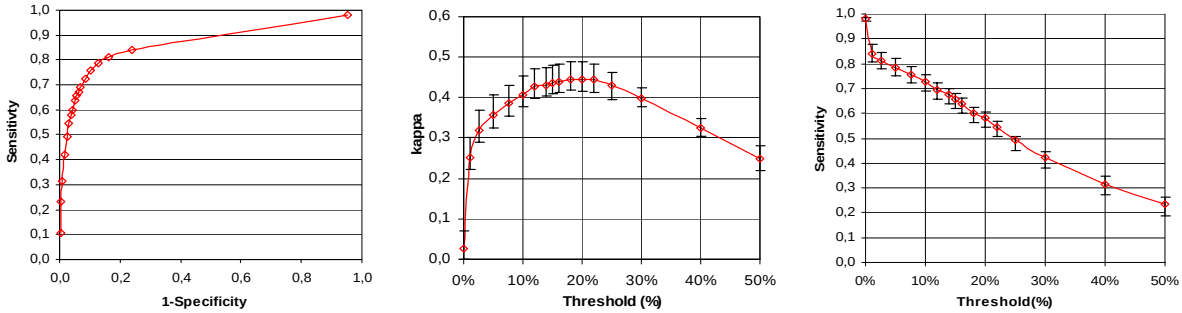
3.5.1. Area calculation

The area affected is the main indicator computed by the automated procedure. From the analytical model it is possible to compute the area affected by drusen and define its contour. Since the model contains only drusen, a threshold over the model will return this information. The threshold was defined in percentage above the normalized background:

$$threshold = (100\% + threshold(\%)) \cdot background \tag{3.36}$$

The threshold value that produces more accurate contours was obtained among a set of 22 images by computing the *false-positives* and *false-negative* pixels between the automated method and trained specialists marked images. From this data the Receiver Operating Characteristic curve (ROC) and Cohen's Kappa statistics were obtained and average values plotted (figure 3.51).

From the Kappa Statistics in figure 3.51.b it can be seen a maximum of *0.446* at 20%, which was visually verified to be corrected. The fact that the sensitivity decreases as the threshold increases could justify the adoption of a lower threshold. However, the visual analysis didn't show that need. Therefore, the more adequate threshold, according to the plots, is a threshold of *20%* above the background intensity, which obtained a *sensitivity = 0.581*, *a specificity=0.962* and a *k=0.446*.

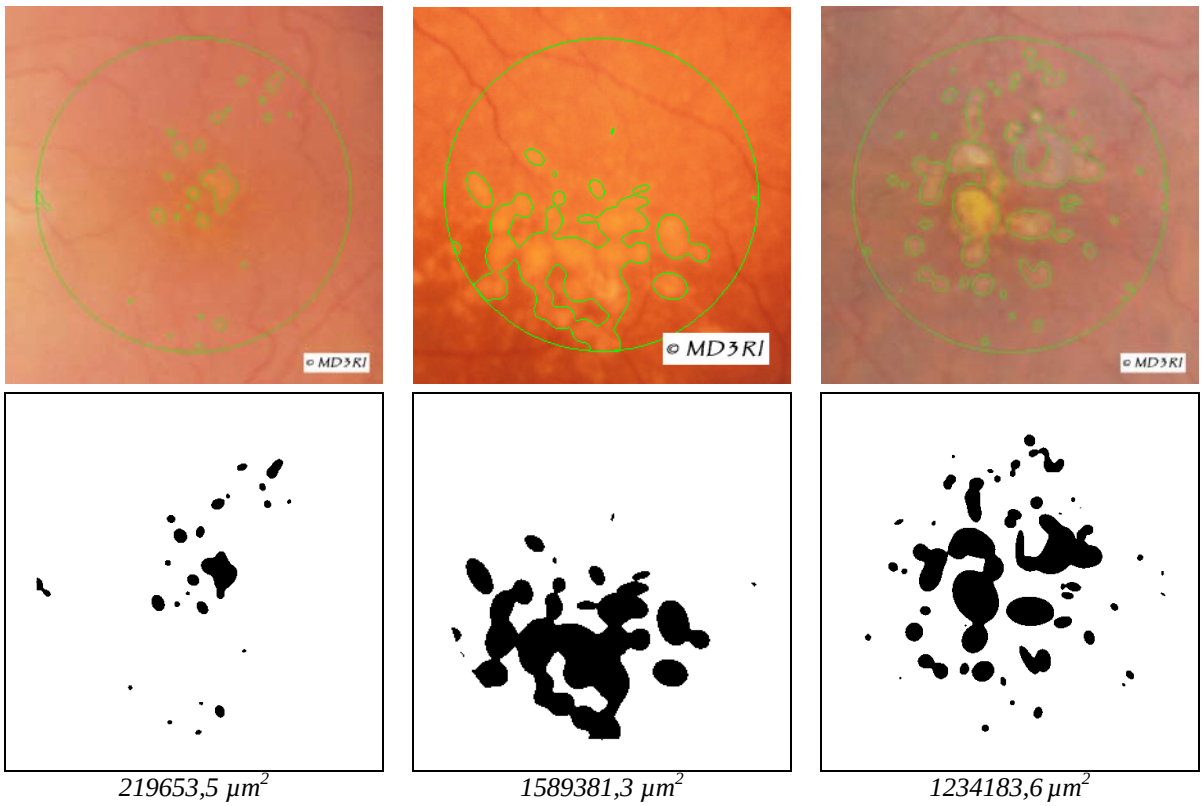


(a) *(b)* *(c)*

Figure 3.51 – Area calculation for different parameterizations. (a) Receiver Operating Characteristic; (b) Cohen's Kappa Statistics; (c) Sensitivity Analysis for images marked by trained specialists.

Figure 3.52 shows some examples of *contour* images and the corresponding *marked* image, which is the model image after the threshold. Although the contours are not as detailed as the images suggest, this modelling approach captures the average contour being more consistent along different analysis.

The final area value is obtained by counting the number of marked pixels in the *marked* image. As different image resolutions are possible the areas' values should be normalized. This is obtained by converting them to square micrometers. As the ROI radius is fixed (*1500 µm*), the total area within the ROI is determined by the area of the close circle ($\pi \cdot r^2$) which can be used to normalize this data.



219653,5 µm² *1589381,3 µm²* *1234183,6 µm²*

Figure 3.52 – Drusen area calculation. (top) contour defined at *20%* of amplitude, (middle) the *marked* image; (bottom) the areas values.

3.5.2. Integral estimation

The genesis of Drusen deposits is an extra-cellular material accumulation between the basal lamina of the retinal pigment epithelium and the inner collagenous layer of Bruch's membrane, what causes the retinal pigment epithelium to be raised with a smooth slope. The materials that cause these elevations reflect more light than a normal retina appearing in fundus images with a brighter and yellowish colour. Using the two-dimensional coordinates - (x,y) - and the image intensity - $I(x,y)$ - from the retinal images it is possible to estimate a tri-dimensional view of its surface where elevation of the spots can be perceived. It should be noticed that it is an estimation and not a real tri-dimensional view since the depth coordinate (intensity) is dependent on illumination.

These observations also suggested the evaluation of the integral for the identified areas. Although this value is dependent on the illumination conditions, with the image normalization, it produces more reliable values. It is an additional indicator useful to study the drusen evolvement in a long term study of the same patient.

The procedure used to quantify the integral (figure 3.53 and figure 3.54) is a discrete method, which sums the individual pixel intensities, after subtracting their background intensity (3.37). It is only calculated for the marked pixels in the area quantification, as in the previous section. The *marked* image, after thresholding the model, will be used as a mask for the integral quantification.

$$Integral = \sum_{x}\sum_{y} i(x,y) - z_o \tag{3.37}$$

To normalize this value for different resolutions it is converted to squared micrometers, as in the previous section.

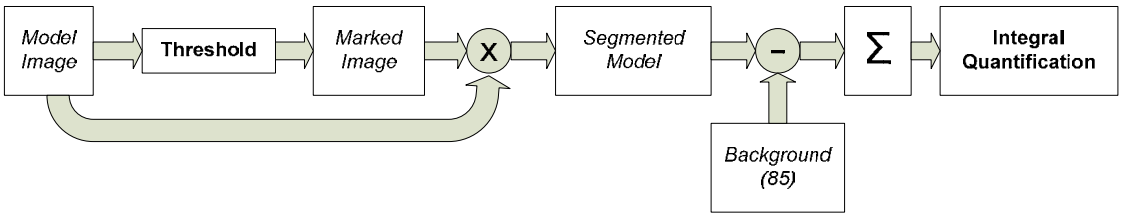


Figure 3.53 – Integral quantification procedure.

3.5.3. Number of Drusen

The number of drusen is used for some epidemiologic studies together with their main class (small, medium, large size). In this methodology this value is obtained directly by the GPL algorithm output, i.e., each RIM detected is considered as a druse. However, this latter is

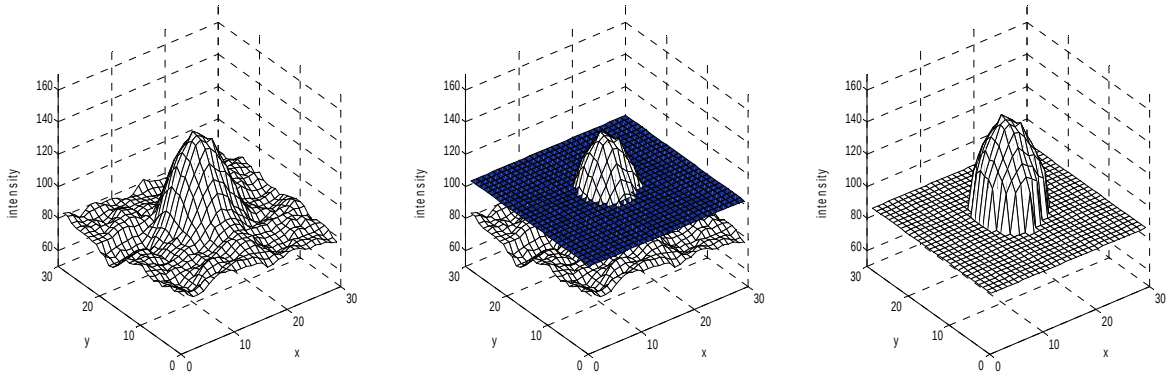


Figure 3.54 – Integral estimation process. (a) Original image; (b) Threshold; (c) Segmented Model.

an approximation, given that the number of GGD functions used in the model is not a reliable measure since it can be detected more than one GGD functions for each real drusen, as showed in figure 3.55. Here it is visible, especially on larger drusen, that there is more than one RIM for each druse.

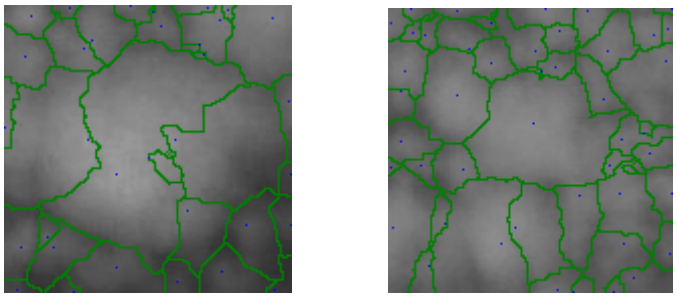


Figure 3.55 – Examples of drusen detection with more than one GGD per drusen (*blue* dots).

3.6 - Summary

The presentation of the methodology for drusen automated quantification is now complete. As key steps in this methodology it should be mentioned the image normalization and the modelling procedures. The *Gradient Path Labelling* is a novel algorithm proposed in this work as a general algorithm for spots segmentation. It is used in this methodology as a preparation step for the modelling procedure.

The image normalization is very important as it prepares the image by standardizing the illumination and contrast for further processing, which can be, in this work, the drusen detection or can be other kind of analysis such as, tumours evaluation, diabetic retinopathy detection or others. However, in some cases the automatic selection of the contrast window fails, requiring the user intervention to select a window which fulfils the requirements (vessels over background). This is mainly due to the image characteristics (deficient lighting, severe abnormalities or other) which can mask the vessels that should be detected by the sliding-

window method. Nevertheless, the benefits for the following methods are compensatory even though the unsupervised procedure is adequate for the majority of the images.

The *Gradient Path Labelling* algorithm showed to be a novel and consistent alternative to other segmentation methods, such as the Watershed Transform. Its ability to detect spots even in low-contrast areas and in images with uneven illumination patterns are one of its main features. Performance is also another of its key features. In the comparison with one of the fastest implementations of Watershed Transform, the GPL was significantly faster and accurate.

The modelling procedure is a step in which it is tried to achieve a shape consistent segmentation. Keeping in mind that drusen exhibit a typical round shape and that confluent drusen are compositions of smaller drusen, this approximation models all these structures to keep the typical drusen shapes in the final segmentation.

With the methodology completely defined, the next chapter will present the software prototypes which implement these algorithms and where they were tuned. In chapter 5 the results of their clinical validation are presented and discussed.

Chapter 4 -

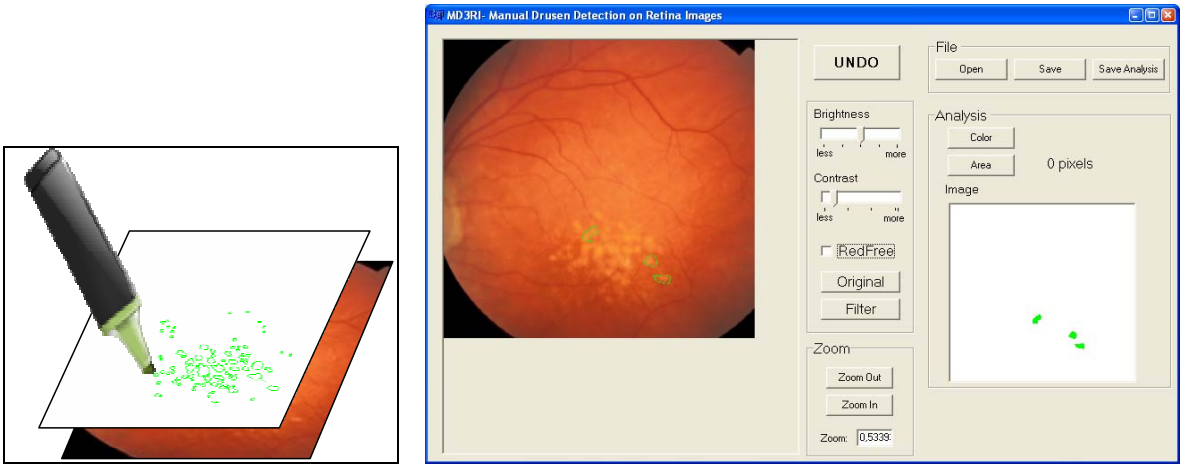
The Methods of Validation

This chapter will be dedicated to the validation of the proposed methodology for drusen quantification. The current Gold Standard method for drusen measurement is the Wisconsin Grading System. It classifies drusen in categories, depending on their shape and size of the bigger drusen found. This method is done manually by trained specialists, in some sense due to the lack of automatic and reliable quantification tools. This fact reduced the number of available sources of images for comparison with the images analysed by the proposed methodology.

The validation of results of the automatic methodology for Drusen detection is made in two phases. In a first phase results will be compared with those produced by clinicians. Since there is a significant inter-operator and intra-operator variability using this procedure it will be impossible to achieve a good accuracy. Therefore, once the main configurations are defined by comparison with clinicians, a second phase will start. Here the results produced by the software will be evaluated by clinicians and parameterization and configurations will be made towards a consensual appreciation.

To create a reference set of images a panel of Ophthalmologists and other trained specialists were invited to mark images, which were then considered the gold standard analysis in this study. They marked the areas affected in the retinal images, which were then used to evaluate the agreement with the automated method.

To mark these images, initial tests were performed to compare the markings done by hand drawing on photographic paper with the ones drawn in computer. Two Ophthalmologists were invited to make this test on two images. In the first procedure, using photographic paper,

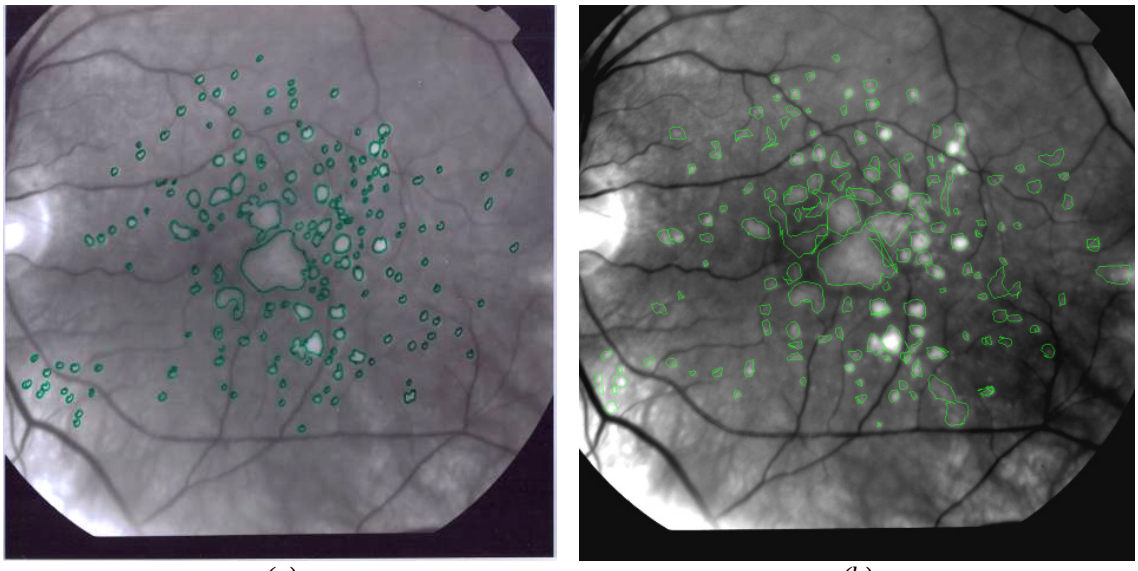


(a) *(b)*

Figure 4.1 – Gold Standard images preparation. (a) hand writing over printed fundus photographs; (b) drawing assisted by computer using a mouse.

the specialists drew the drusen contours on a clear transparency placed over the printed image (figure 4.1.a). The transparency was afterwards digitalized and the contours filled with a solid pattern. To draw the contours in the computer a specific program was developed to draw closed contours using a mouse (figure 4.1.b). This program allows to undo actions and to easily show/hide the contours being drawn to provide better perception of their quality. In figure 4.2 is shown one of the test images marked by hand and by computer.

Although the accuracy of the contours were better in the manual drawing over transparencies rather than using the computer, the flexibility showed by this second method proved to have more acceptance by the specialists. Here, they can easily correct contours by undoing actions and by adjusting zoom, brightness, contrast and others, making this their



(a) *(b)*

Figure 4.2 – Example of one image marked by hand (a) and by computer (b).

preferable choice. However a semi-automatic contour detection feature was suggested to simplify the drawing of contours.

This software program was further developed to include this and other functions to assist the specialist in the drawing of the contours, being the tool used by the specialists to mark the images. It now allows drawing drusen contours in a semi-automatic procedure.

The validation of the drusen automatic detection methodology involved the development of another independent tool that implements the methods described in chapter 3.

These two contour detection tools were denominated *Automatic Drusen Deposits Detection in Retinal Images* (AD3RI) and *Manual Drusen Deposits Detection in Retinal Images* (MD3RI) for the automated and manual detection of drusen, respectively. These tools will be presented in the following sections.

4.1 - Manual Drusen Deposits Detection in Retinal Images

The MD3RI is a user-friendly application developed for Ophthalmologists to analyze and mark the presence of Drusen deposits in retinal images. The results are produced in a suitable format for comparison and areas calculations. It was developed using Borland C++ Builder v6 (2002), by being a *Rapid Application Development* environment. The tool was published by Mora, Vieira *et al.* (2006) and published online (Uninova 2005).

This tool was developed with the purpose of evaluating the performance of the automatic Drusen analysis methodology described in Chapter 3 by comparing its results with

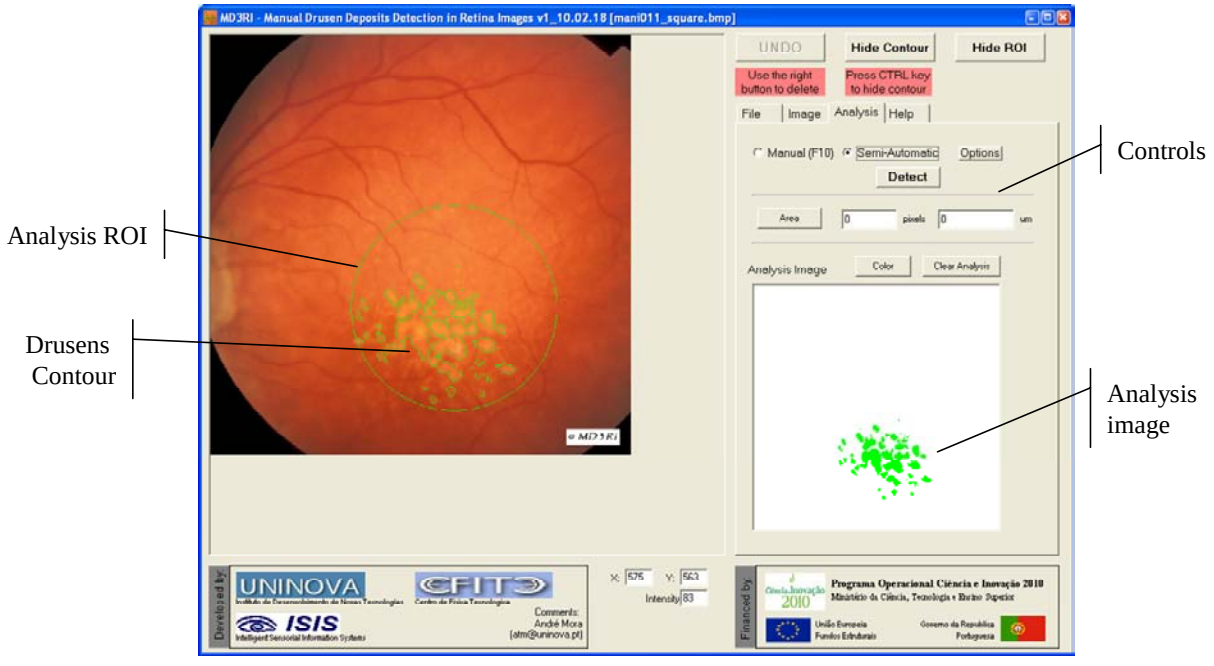


Figure 4.3 – Example Screenshot of MD3RI User Interface.

the same images analyzed by Ophthalmologists in different Medical and Research Centres. The contribution of different Centres around Europe allowed a greater diversity of analysis criteria, resulting in a statistically more precise comparison.

Marking drusen on retinal images is a difficult task to reproduce, added to the fact that external factors such as ambient illumination and different screen technologies (CRT, LCD and TFT) and adjustments affect the way the user visualizes the image. This however, can be minimized by using hardware monitor calibrators or software calibration images (figure 4.4). The recommended use for MD3RI is to use calibration images for adjusting illumination and contrast settings, in order to clearly identify the full gray scale levels.

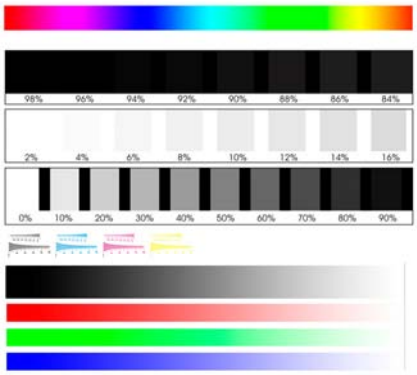


Figure 4.4 – Example of a calibration image (taken from (High Display 2010)).

The program's main functionalities include:

- Display and Save Marked Image;
- Store and load analysis from file;
- Region of Interest configuration, store and load from file;
- Image enhancements;
- Manually and Semi-Automatic contour drawing tools.

The region of interest (ROI) that ophthalmologists considered to be significant for drusen analysis is a circle with approximately twice the diameter of the optic disk centred in the macula (figure 4.3). These dimensions followed the recommendation of the Wisconsin ARMD Grading System (Klein, Davis et al. 1991). After defined this circle, the software only allows contours to be drawn inside this area. The ROI definition can be loaded from or saved to a file, so that the analysis configuration can be easily reproduced on other computers participating in the same trial.

For better image visualization, especially on images with non-uniform illumination, a set of simple image enhancement tools were added (figure 4.5). The first option is the colour mode, which can be: original colour, red free (green + blue colour channels), green colour

channel or grey scale. The red-free and the green channel colour modes are commonly used in this type of images to improve image contrast. Image brightness, contrast and equalization adjustments were also added to improve users' visual perception. Mean filtering for image noise reduction is available, contributing to smoother and better defined contours. Zoom capabilities were also added to improve the usability, as well as a tool to compare the size of a druse with a reference circle (63µm, 125 µm, 250 µm or custom).

The method used by this program is based on two images (figure 4.6); the *contour image* that is visible to the user, where the contours are displayed and drawn, and the *marked image* that is a working image with the same size and where the analysis is stored. This latter is a white background image where the affected areas are marked in colour, as shown in figure 4.8.b, while the *contour image* is obtained by the contours of the marked areas (figure 4.8.c) combined with the original image (figure 4.8.d).

This method has the advantage of being able to retouch the contours only by clearing the area in *marked image* and by allowing the manual and the semi-automatic drawing procedures to coexist. Also, the analysis can be stored and loaded from a file at any time and easily compute the affected areas. For drawing contours, two modes are available: the manual and semi-automatic.

In manual mode the contour is drawn pressing the mouse left button and moving it around the drusen contour. When the user releases the mouse button, the line is automatically closed and it draws on the *marked image* the same shape but filled. At this point, the Drusen

Figure 4.5 – Image enhancement panel.

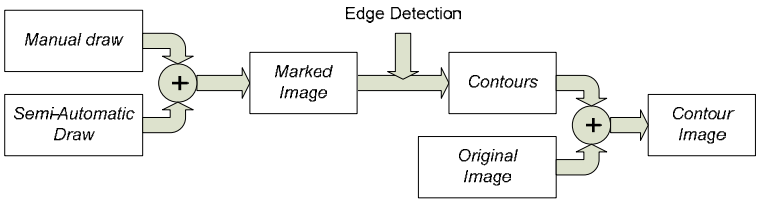


Figure 4.6 – MD3RI functional drawing procedure.

contours of the *contour image* are updated with the contours of the shapes drawn. To retouch the contours this operation can be repeated to add new areas or by using the mouse right button to remove areas.

In semi-automatic mode it is used one of the outputs from the GPL algorithm described in section 3.3.3, the drusen area of influence. In this drawing mode the user selects with the mouse left button a druse and by moving it to the left or to the right the druse contour will automatically grow or shrink. The principle is to add to the *marked image* the pixels belonging to its area of influence that are above a certain threshold, which is adjusted by moving the mouse cursor to the left or right in order to lower or to increase it, as shown in figure 4.9. In this mode the mouse right button, simply removes the druse area from the *marked image*. For initializing the structure containing the located drusen and their area of influence it is needed to press the *Detect* button (figure 4.7).

When comparing manual and semi-automatic drawing modes, the choice fell on the latter due to its easier operation and better defined contours. In some cases, when the semi-automatic produces a contour that needs to be slightly refined, the manual operation can always be used as a complement.

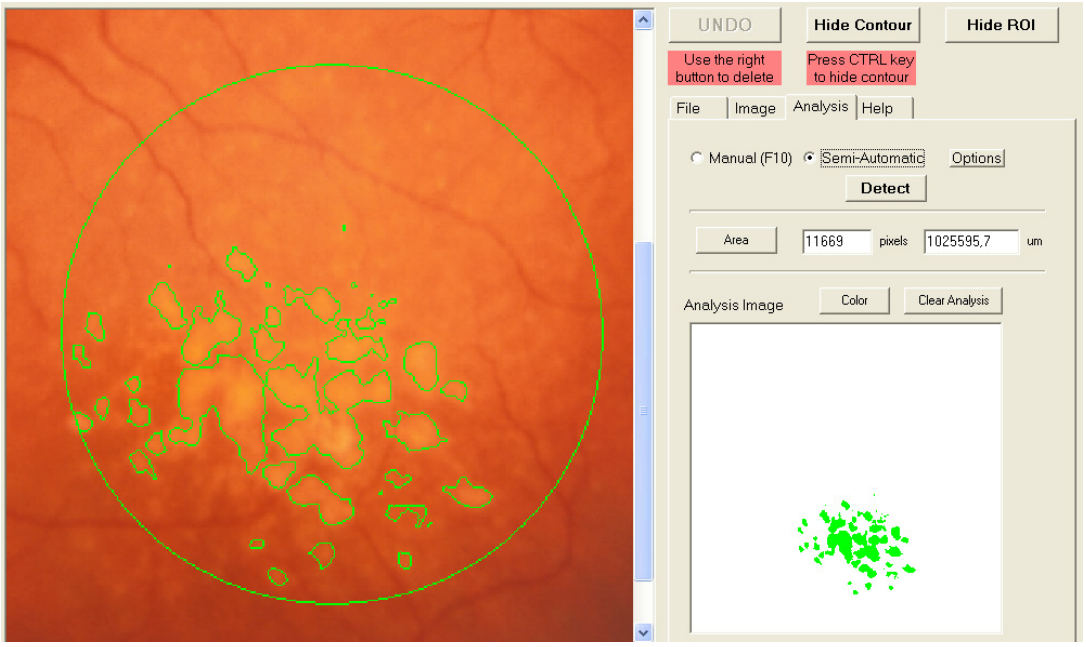


Figure 4.7 – Drawing mode configuration and area calculation.

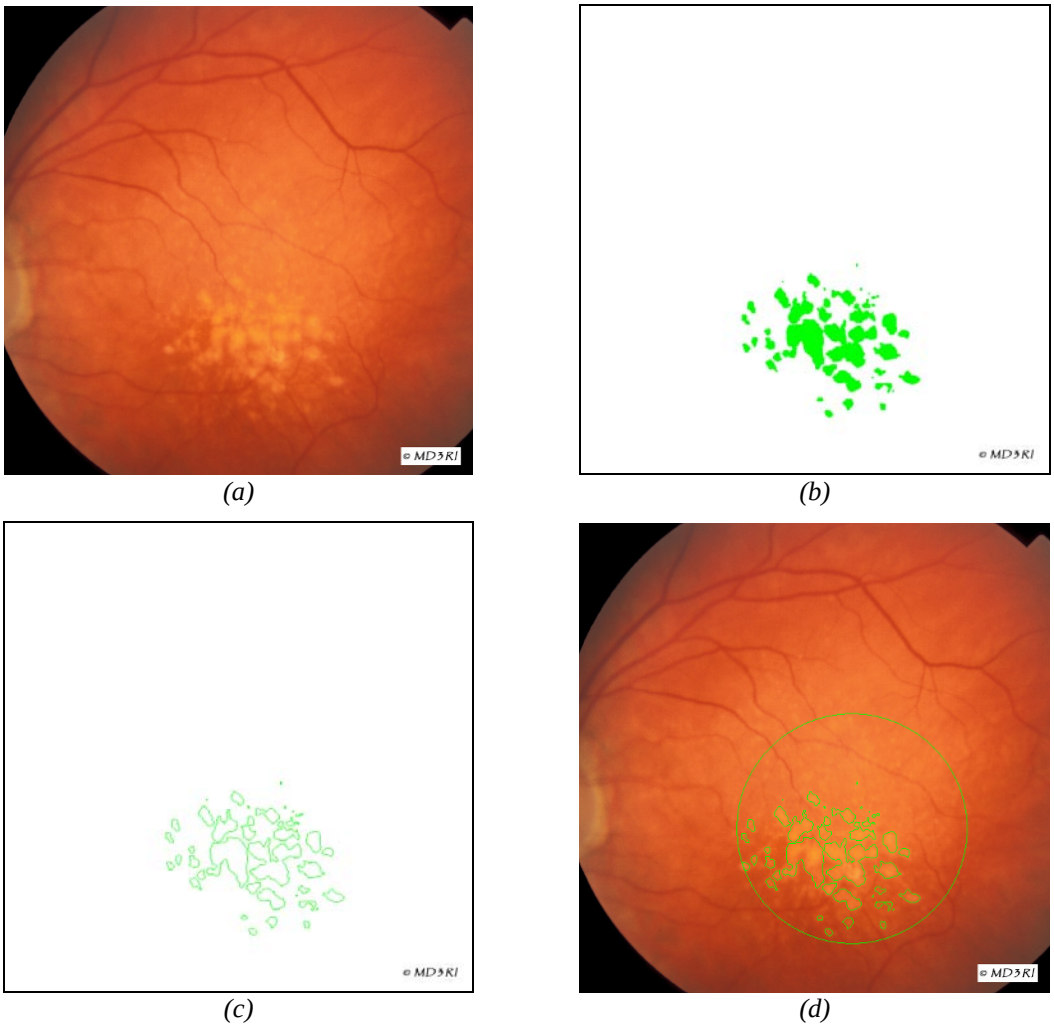


Figure 4.8 – Drusen contour using MD3RI. (a) original image; (b) *marked image*; (c) contours; (d) *contour image*.

During the analysis and for better perception of the next contour, users can temporarily hide the current contours and the ROI circle. This can be done using the buttons *Hide Contour* and *Hide* ROI or pressing the CTRL key.

After marking all the drusen it is possible to compute the total area of drusen drawn. This area is expressed in pixels and in square micrometers. To calculate it in square micrometers it is considered that the ROI has a fixed radius of 1500 μm.

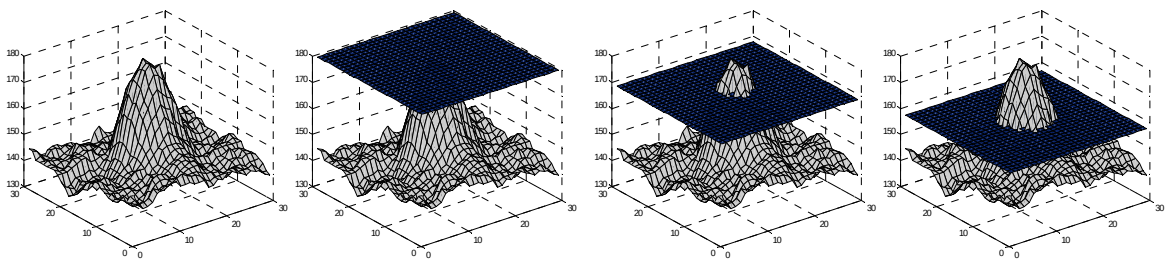


Figure 4.9 – Example of MD3RI semi-automatic contour drawing procedure.

4.2 - Automatic Drusen Deposits Detection in Retinal Images

The AD3RI is a C++ application developed for the automatic quantification of drusen spots in retinal images. The algorithms were in the beginning of this work developed under Matlab environment (MathWorks 2002) and were later re-implemented in Borland C++ Builder to achieve a better performance. The reduction in time consumption was significant; it was reduced approximately by a factor of 200. The sectioning and the modelling algorithms were based on the implementation of Moitinho (2009) with improvements on performance and accuracy. Over his implementation it was introduced parameter constraints and other functionalities to improve the algorithm performance.

The main purpose of the program is to be an image processing tool, with specific features for drusen quantification. This includes all the necessary steps to model and quantify drusen. Its main functionalities involving the quantification of drusen are:

- Retinal images background correction;
- Region of interest definition;
- Automatic and manual drusen detection
- Drusen modelling;
- 2D and 3D image visualizations; and
- Areas and Volumes calculation.

The drusen functions are all condensed in one multi-panel window as shown in figure 4.11.a. These panels represent the several steps of the quantification methodology. In the first panel, a list of steps for drusen quantification and some pre-processing steps for image enhancement (green channel selection and mean filtering) are shown to the user. The checkboxes on the left of the step buttons allow the user to configure which step to run when doing a Fully Automated operation.

The functionality *Crop to Centre* is intended to centre the retina image and clear the dark borders that usually appear in fundus images. The algorithm to implement this operation is represented in figure 4.10. It applies the negative, a threshold and a connected components algorithm to isolate the bigger object in the image. Then, knowing its centre it calculates the bigger square region which sizes inside it and crops the image accordingly.

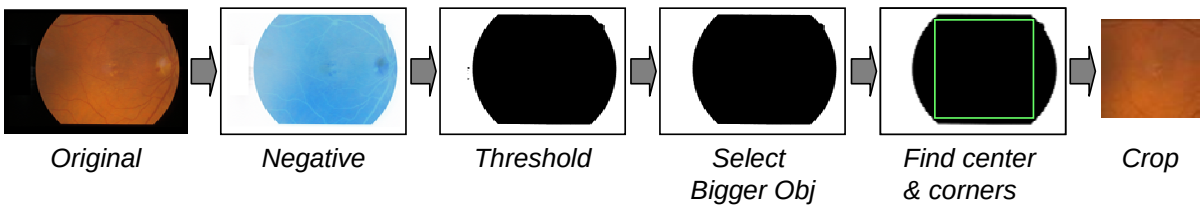
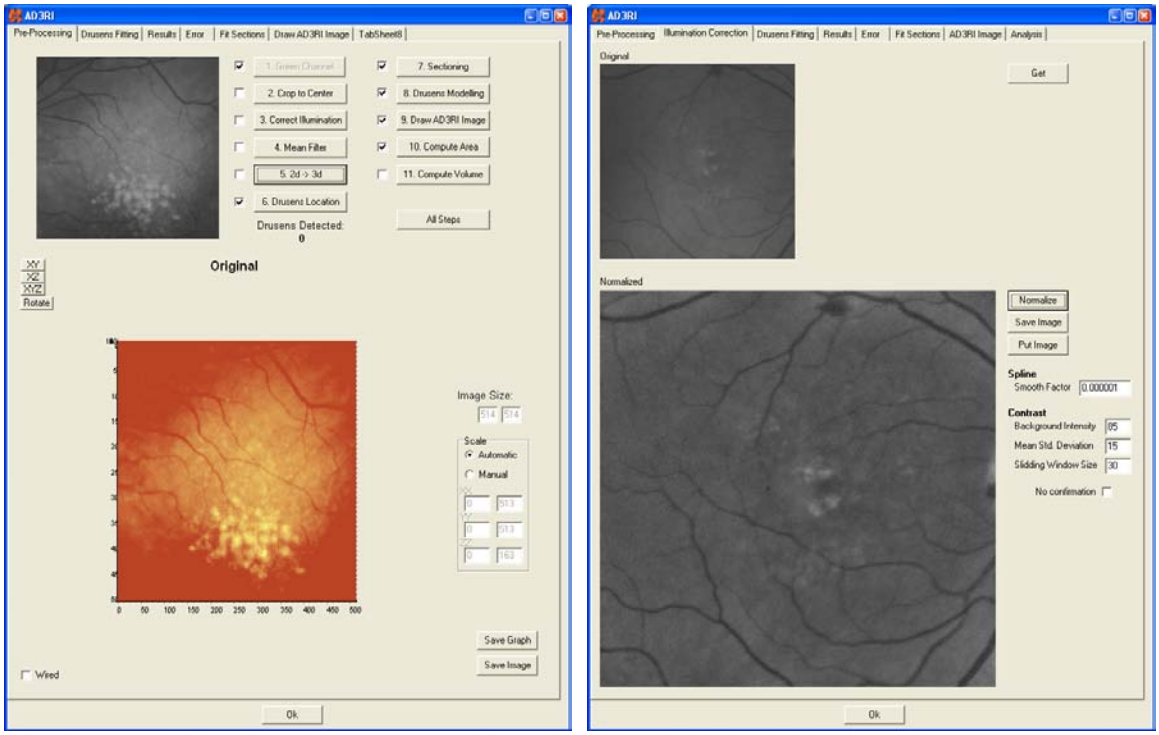


Figure 4.10 – Algorithm of operation *Crop to Centre.*

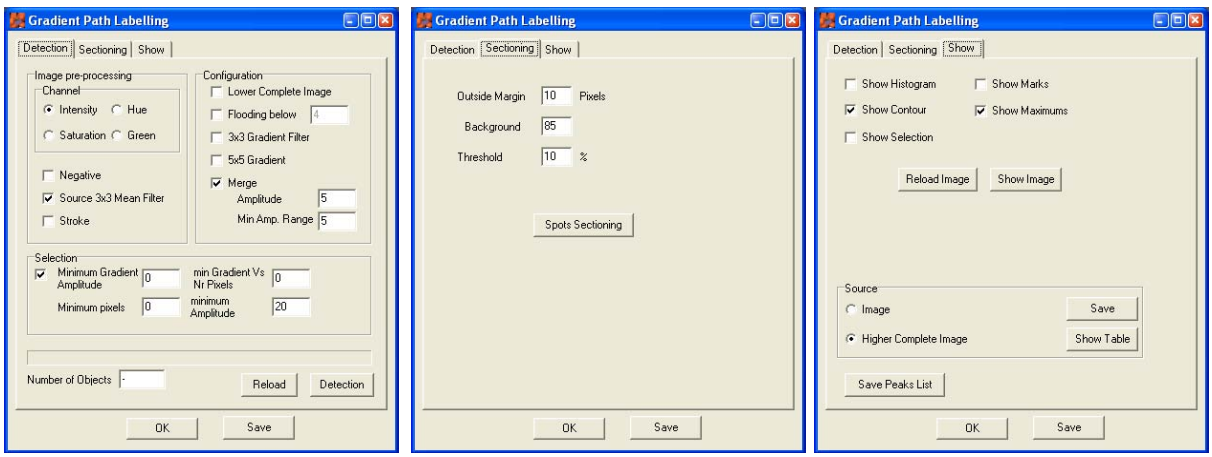


(a) *(b)*

Figure 4.11 – Drusen quantification window. (*a*) Entry panel; (*b*) Image normalization.

The Illumination Correction panel (figure 4.11.b) is where the image normalization is presented. This step was implemented in the Matlab environment and was exported to an executable file using the Matlab Compiler Tool. The reason for the Matlab implementation was due to the unavailability of free software libraries containing the Smoothing Spline fitting implementation. The only disadvantage for the user is a slower execution, since it loads the Matlab Runtime Environment for this operation. In this panel it is possible to configure the parameters for the smoothing spline fitting and the contrast normalization.

The drusen location window (figure 4.12) is the frontend from the Gradient Path Labelling algorithm and the sectioning algorithm. The first panel -*Detection*- configures the



(a) *(b)* *(c)*

Figure 4.12 – GPL and sectioning frontend. (*a*) detection; (*b*) sectioning; (*c*) results presentation.

several steps of the GPL algorithm: the image pre-processing; the labelling procedure; the spots merge and spots selection. The second panel -*Sectioning*- configures and executes the *sectioning* algorithm. The result from the *Spots Sectioning* are a sequence of images named with a postfix *IDS* and a file containing a description of each of them (name and location) and the configurations used (margin and background) so that the original image can be reconstructed. The last panel -*Show*- is used to visualize the results from the GPL algorithm and save auxiliary data. The system allows showing (figure 4.13): the location of the detected maximums, their contour, their areas of influence and the pre-sectioning areas.

The *Drusen Fitting* panel (figure 4.14) contains the interface for the modelling kernel, which is based on the Levenberg-Marquardt algorithm. The implementation of this algorithm was adapted from the Numerical Recipes in C (Press, Teukolsky et al. 1992) and proved to be robust and fast. Some modifications were introduced to allow the definition of parameters' constraints. In this panel it is possible to configure and test the modelling procedure. It allows the configuration of the model to be fitted by adding and removing GGD functions and to configure their parameters. Three ways to add functions to the model were introduced: by manually typing the function parameters; by clicking over a thumbnail image which adds a function pre-configured to the location and amplitude of the selected pixel; and by using the location and amplitude of the spots detected by the GPL algorithm. In order to allow a deferred processing, the functionalities of storing and loading from a *CVS* file were added also.

During the fitting operation the chi-square error is shown graphically to the user, as well as the evolvement of the damping parameter -λ- (lambda) and the number of iterations. The fitting operation can be stopped at any time and can be re-run unlimited times. At the end of the fitting process the new model is drawn in a chart, which can be viewed in 2D or 3D as shown in figure 4.14.

(a) *(b)* *(c)*

Figure 4.13 – Output images from the Gradient Path Labelling. (a) contour (*green*) and maximums (*blue*); (b) areas of influence and maximums (*blue*); (c) pre-sectioning areas.

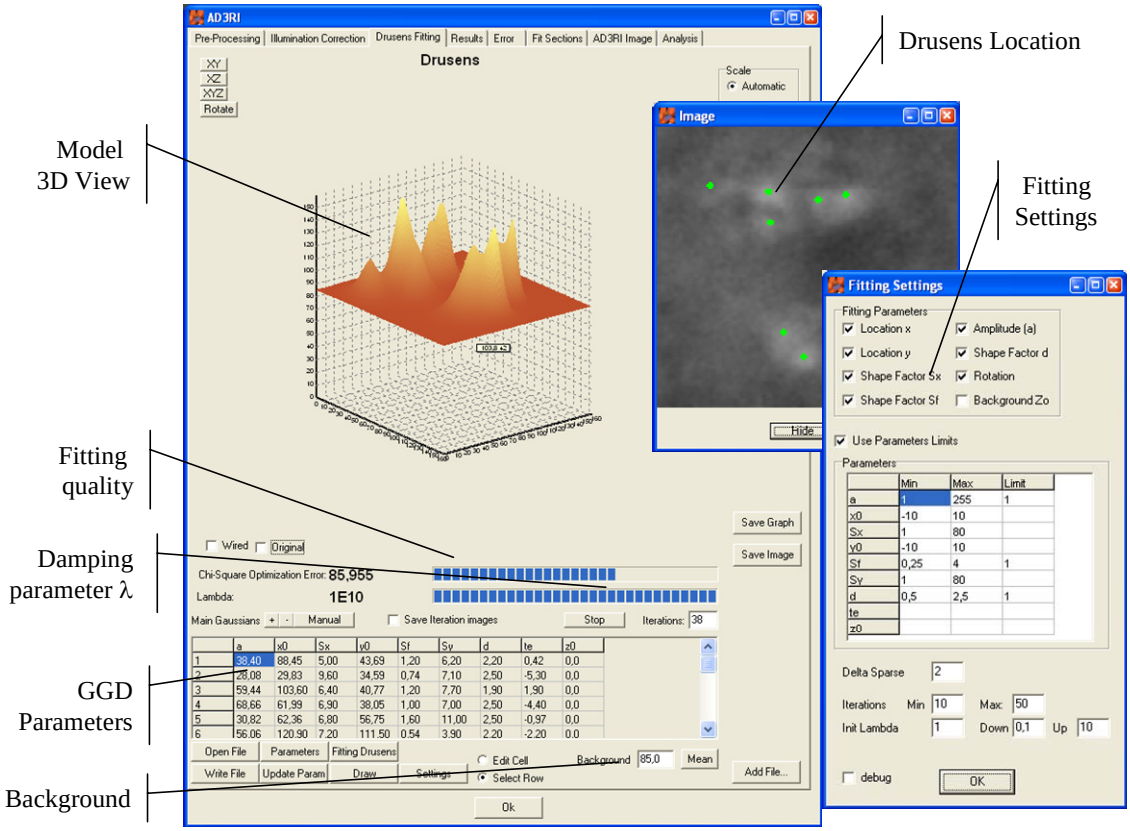
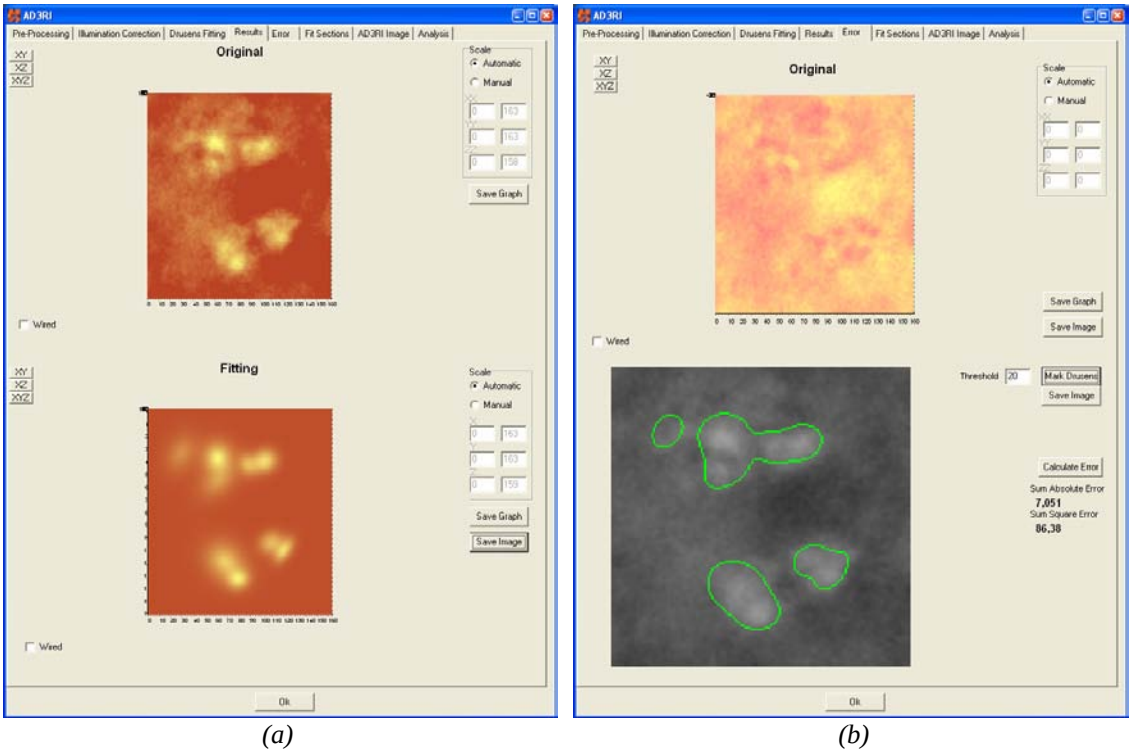


Figure 4.14 – Panel for the drusen modelling.

The *fitting setting* window shown on the right side of figure 4.14 configures the Levenberg-Marquardt algorithm. In this window the user can specify the GGD parameters to be adjusted during the fitting; configure the parameters maximum and minimum limits or no



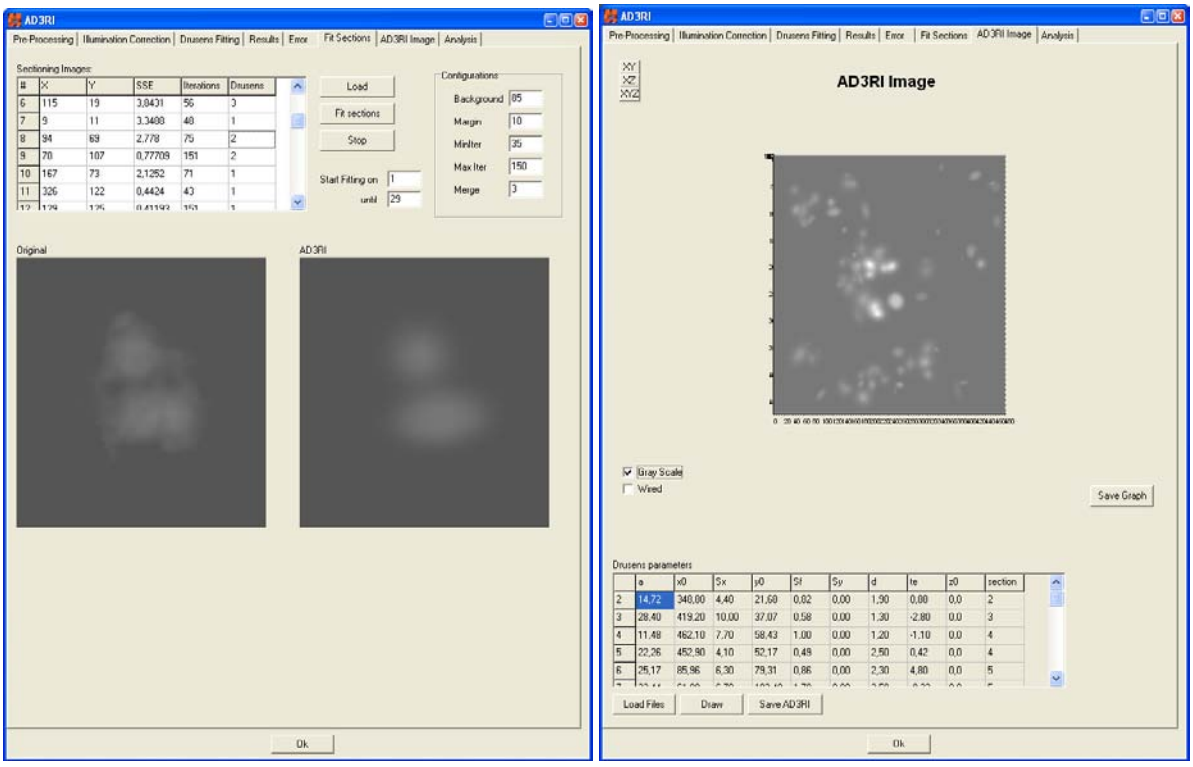
(a) *(b)*

Figure 4.15 – Modelling results panels. (a) Original vs. model; (b) Error and contour.

limits; the sparse interval; the stopping and the evolvement conditions for the Levenberg-Marquardt algorithm.

After the modelling procedure the user can analyse the produced data in the *Results* and *Error* panels (figure 4.15). In the first panel the model can be visually compared to the original image in two-dimensional or three-dimensional views as well as in colour or grey scale. In the *Error* panel the user can observe a chart containing the differences between the model and the original image. On the bottom of this panel the user can evaluate the contours of the model by thresholding it to a defined percentage above the background intensity.

The panel *Fit Sections* (figure 4.16.a) is dedicated to automate the processing of the sectioned images, which contains a reduced number of drusen. It reads the file containing names and locations of the sectioned images, as well as their configurations, and displays them in the interface. The *Fit Sections* button is responsible for applying the drusen detection and the drusen modelling to each of the *section* images, starting and ending with the numbers defined on the interface. Finally, the model image is saved with its original filename and a postfix AD3RI. During this process the table *Sectioning Images* is updated with quality information from the modelling, namely the SSE, the number of iterations taken and the number of drusen. At the end the user can compare the generated model with the original



(a) (b)

Figure 4.16 – Modelling using the *sectioning* procedure. (a) fitting the individual sections; (b) compute the overall model.

image by selecting a row from the table; this operation will display both images side-by-side for visual inspection.

The next step is to gather the individual sections into one single image, what can be executed in the AD3RI *Image* panel (figure 4.16.b). The function parameters that were obtained in the previous step are loaded into a table containing all the functions from the section images. With this information the full model is calculated, saved and presented on a chart which can be analysed in two-dimensional or three-dimensional views.

At the panel *Analysis* (figure 4.17.a) the drusen are quantified and the images that can support this analysis are produced, namely an image containing drusen contours and other containing the *marked* image with the affected areas. This latter is identical to the *marked* images produced by MD3RI, which allows their direct comparison.

In this panel the user can configure the background value and the threshold that separates drusen from background. After pressing the A*nalysis* button, which generates the *contour* and *marked* images, the user can calculate the areas in percentage, the areas in square micrometers or estimate the absolute volume in cubic micrometers (the intensity was considered to be in the same units). These indicators were calculated in five different areas

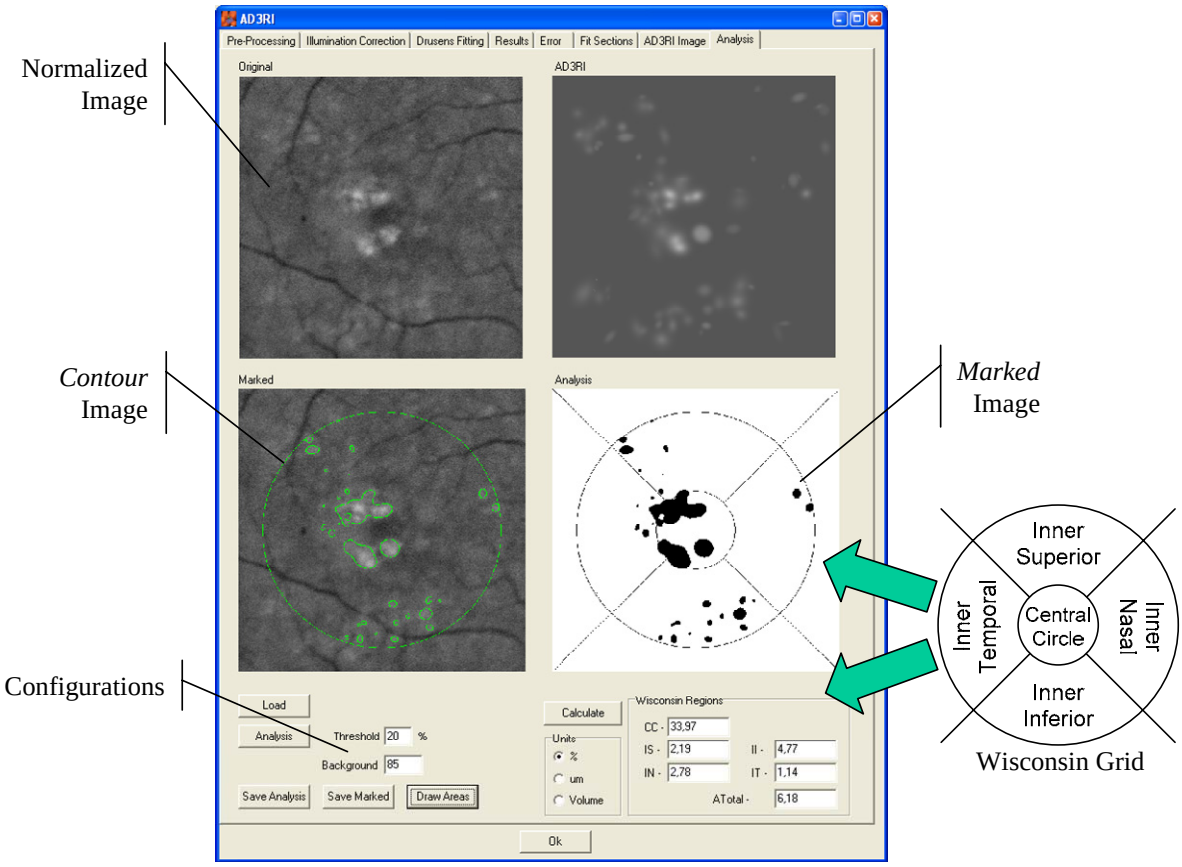


Figure 4.17 – Quantification of drusen and marked Images

(*CC*, *IS*, *II*, *IN*, *IT*) according to the recommendations by the Wisconsin ARMD Grading System and also their total area (*ATotal*).

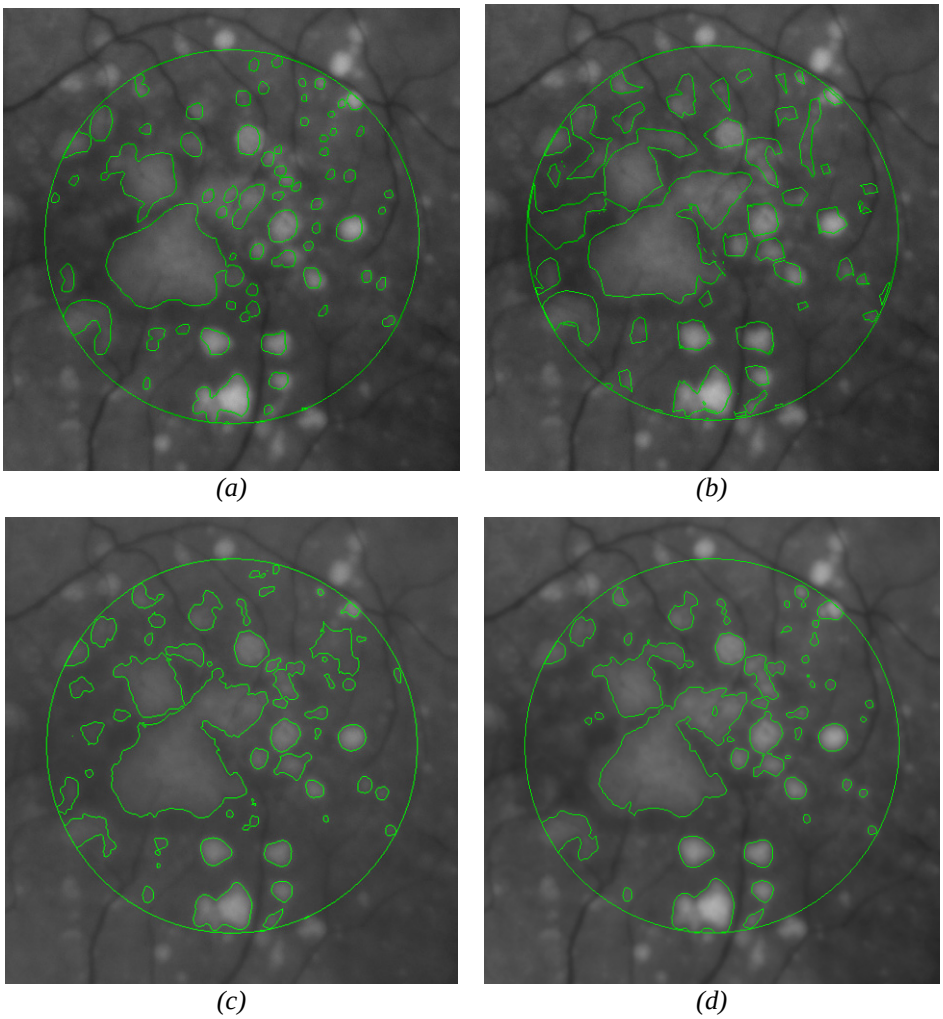
The methodology for drusen quantification proposed in this thesis ends at this step. Although the description just presented is very detailed, the analysis can be fully automated and made in just one button press. This detailed process was needed to debug and to parameterize the methodology steps, but since the default configurations were defined, it can run automatically. All the parameters, except the ROI definition ((x,y) location and radius), are stored in configuration files or are extracted from the image, which frees the user from configurations during a normal operation.

To process images datasets, it was added a functionality to AD3RI software to automate their analysis. It receives the list of images to analyse and the automatic quantification is executed using the default parameterization.

4.3 - Summary

Both software have successfully achieved their objectives. The MD3RI was designed to be used by the specialists involved in this study to analyse our images datasets, which was completely accomplished. The specialists that were involved, a panel of Ophthalmologists from Portugal, Scotland, Brazil and Austria and of Biomedical Engineers, gave a positive feedback regarding its usability and accuracy. Actually some of the specialists use the software as the analysis tool in their own clinical studies. The validation of the AD3RI software will be done in the next chapter, with an extensive comparison with the markings performed by the specialists.

The use of semi-automated marking of drusen in MD3RI increased the accuracy and decreased the time spent on the task itself. The manual marking over photographic paper has several alignment problems and lacks usability. Also the marking on basic image drawing software does not guarantee the accuracy needed as it is very difficult to keep the mouse stable during the contour definition. Figure 4.18 shows an example used in the comparison among these several techniques. The two problems mentioned (alignment and irregular contours) are clearly visible in figures (*a*) and (*b*). The images (c) and (d) show the same image marked using the MD3RI and is clear that these problems were overcome. The major problem is still the variability among the specialists. Figures a, b and c were analysed by the same ophthalmologist in different occasions and it is clearly visible the difference in the detection criteria among the different analyses. The same difference was observed among different Ophthalmologists.



(a) *(b)*

(c) *(d)*

Figure 4.18 –Drusen contours drawn with MD3RI. (a) drawn on photographic paper (b) MD3RI in manual drawing mode; (c) (d) MD3RI in semi-automatic drawing performed by two ophthalmologists.

The AD3RI software implemented several new image processing algorithms for the automated quantification of drusen. The paradigm of modelling the images intensity was one of those which allow the creation of a shape controlled segmentation. This can be useful when the contours are irregular, diffuse and difficult to define. This approach is able to model the mass of the spot that contributes to a more reproducible segmentation method. It keeps the typical shape for the objects that are being modelled overcoming some undesired effects due to deficiencies in illumination, scratches, low resolution and others.

The quantification process was fully automated in the AD3RI application. The default configurations were successfully defined by comparison of the images marked by the specialists with the ones produced by the software. These were obtained after applying sensitivity tests to choose the best parameters that produce similar markings to the specialists.

Chapter 5 -

Results Evaluation

The assessment of the accuracy of the automated quantification of drusen is in the first phase done by comparison with the markings done by specialists. On a second phase the specialists were asked to comment on the results produced by the automated process. For the first phase the objective is to approximate the result as must as possible to the ones produced by the specialists, while in the second phase the goal is to fine tune the algorithm.

The study involved various independent specialists in order to construct a ground truth to which the comparisons could be made. However, the variability that is observed within specialists and among specialists made it difficult to compare between them and the automated method. Thus the creation of a ground truth analysis was considered to be not completely feasible due to the mentioned variability. As consequence it was decided to compare the specialists among themselves to evaluate their accuracy compared to the others. The objective for the automated method was then to be at least as accurate as the specialists were among themselves, which was considered to be a valid assessment.

In this work it was assessed the accuracy in detecting the affected areas and the repeatability (Bland and Altman 1986). Total values for the indicators and pixel-to-pixel matching were the measures performed for the comparison between manual and automated detection. The repeatability of the automated method for images taken in different occasions is important to analyse the progression of drusen areas. In this work this parameter was also assessed by using repeated images taken at different occasions and the images from the ECAM study (Brunner, Krebs et al. 2005). This study evaluates the progression of ARMD six months after surgical intervention to cataracts. The researchers involved are trying to prove

that ARMD is not changed by this surgical intervention. Therefore, the images before and after the intervention should exhibit approximately the same areas affected by ARMD. The repeatability of the automated method was evaluated considering this hypothesis to be true.

There are two publicly available retinal images database: the DRIVE (Staal, Abràmoff et al. 2004) and the STARE (STARE 2000).

The DRIVE database has been established to support projects on retinal vessels segmentation. It is located on a website where the images are stored and where researchers all over the world can submit their own retinal vessels segmentation results for direct comparison with other algorithms and with the gold standard images. It contains 40 images obtained from a diabetic retinopathy screening program, in which 33 do not contain any pathological signs and 7 contain early diabetic retinopathy signs. The absence of images showing drusen invalidated the interest in this database for the software assessment.

The STARE database (STructured Analysis of the Retina) on the other hand was developed to support more than one project related to retinal image analysis. It contains a total of 402 retinal images, from which 116 were categorized to be used for Drusen Detection Using Multi-level Analysis (Brandon and Hoover 2003) and separated into six types (*normal* (38), *large many* (20), *large few* (12), *fine many* (10), *fine few* (17), *abnormal* (19)). This database contains also the results from the authors' automatic classification algorithm, described in section 2.1.2.5. This DRIVE dataset however was also not considered valid for testing the developed software, since their algorithm does not perform drusen quantification.

5.1 - Materials & Methods

To assess the accuracy of the automated method the images from four different datasets were used and involved a total of nine Ophthalmologists (A.Fa., A.Fo., C.N., D.L., S.R., J.G.O., R.R., T.F. and S.B.) and four trained graders (J.F., P.V., A.M. and F.M.). Due to a lack of availability of some of these specialists, they didn't participate in all the studies of the software assessment, as it will be seen later on.

The assessment was divided into four studies: two for evaluating the software in comparison with the specialists (*accuracy*) (studies: *TRIAL I, TRIAL II*) and two to evaluate the software *repeatability* (study: *REP I, REP II*).

5.1.1. Images Datasets

The images used in this study came from different sources, which were kindly made available by ophthalmologists working in different Clinical Centres. The resolution and

quality of the retinal images were all within the recommended levels. Approximately fifty of them were sent in slide and were locally digitalized. The images datasets included:

- 99 red-free images taken at University of Aberdeen
  - 30º and 35º
  - 1024x1024 pixels
- 24 colour images taken at University of Aberdeen
  - 40º
  - 1024x650 pixels
- 77 colour images from ECAM study in Rudolf Foundation Hospital
  - 40º
  - 1639x1116 pixels

For the *accuracy* studies two sets of images were created from the three first datasets. Only the images which were centred on the macula and that had acceptable acquiring conditions, were considered valid for this work. Especially on the ECAM dataset there were several images which detail was not acceptable due to the opacity of the eye lens and as consequence these were left out from the dataset.

The images were all saved with 256 intensity levels for each colour channel. The retinal slides were digitalized using a slide scanner (Coolscan LS-2000, Nikon Corporation, Tokyo, Japan) and were saved using the same definitions as the other images.

Before distributing the images to the specialists for marking, the ROIs were initially placed in their correct location. As it could be a potential focus of variability, this guaranteed that the images could be compared directly. The *field-of-view* of the images varied among the datasets and even within the same dataset. This required that the optic disc diameter had to be measured for every image to define the ROI width, which is twice its value.

For the *repeatability* study it is required that the ROI of two images are in the exact same location. For that reason, the ROI was defined for one of the images (*source*) and the other image (*target*) was registered to the source. It was used a linear registration process which distorts the *target* image using a combination of translation, scaling and rotation. This operation required the manual location of at least 3 control points in both images. In this study 4 control points were actually located circularly around the macula; these were vessels bifurcations or well defined vessel corners. This operation was done using the Matlab's *Image Processing Toolbox* (MathWorks 2002).

Following the medical and ethical recommendations before sending the images, they were also anonymized by manually removing any patient identification.

5.1.2. Methods

A comparative analysis of the images can be performed by comparing their *marked* images. This is done assessing and comparing overall values and also their local agreement. Although the total area is the indicator that the specialists use to evaluate the size and progression of drusen, it may not be sufficient to reveal if there is agreement between two analyses. For example, the same total area can belong to completely different locations. To overcome this blind effect, agreement indicators must also be used to compare total area and the location of the affected areas. Therefore, for the comparative analysis both total values and pixel-to-pixel analysis were computed. These included the total affected area calculation, the assessment of correlation on total values between specialists and software and other pixel-to-pixel analysis such as, the *sensitivity* and *specificity* analysis or *kappa* statistics. Due to the variability observed among the specialists analysis it was developed a method for estimating a ground-truth analysis among the specialists and assessing the agreement between this analysis and the automated method, which was named Weighed Matching Analysis.

For the comparative analysis of the images the Wisconsin grading circle was used dividing the macula into 5 subfields, as shown in figure 5.1. This allowed the evaluation of the uniformity of the detection criteria along the whole macular area and provides the ophthalmologists regional information of drusen to understand its progression. The Inner Temporal and Inner Nasal are located on the left or on the right depending if it is a left or right eye. A retina image from a left eye has the optic disc on its left and consequently the Inner Nasal on its left too, the opposite occurs with the right eye.

To obtain the indicators values it was used the software AD3RI, Matlab (MathWorks 2002) and SPSS (SPSS-Inc. 2008).

5.1.2.1 Areas Analysis

The area calculation using the method explained in section 3.5.1, is obtained from the *marked* image by computing the number of pixels marked (i.e., not belonging to the

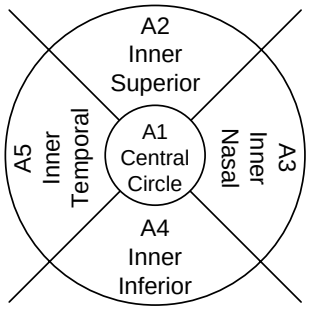


Figure 5.1 – Wisconsin grading subfields (right eye).

background). It is obtained for the whole ROI and obtained also for each of the Wisconsin subfields. These values are then converted into micrometers in order to be comparable between images with different resolutions.

To assess the agreement between different markings (automated and manual), for each image is it calculated the mean, standard deviation and *coefficient of variation* (CV) (5.1) of the different measurements. The latter was also used to identify the analysis outliers, as it will be seen later on each study.

$$Coefficient\ \ of\ \ Variation = \frac{standard\ deviation}{mean} x100\% \qquad (5.1)$$

5.1.2.2 Intra-Class Correlation Coefficient

In order to evaluate the overall agreement in the drusen areas between the specialists and the automated method, it was calculated the *Intra-Class Correlation Coefficient* (ICC) (Fisher 1970), which is usually used to measure the inter and intra-rater agreement. The ICC is a type of correlation which is capable to evaluate the degree of relatedness between different measurements of the same quantity by different observers as groups and not only as paired observations.

Since the original publication of the ICC (Fisher 1970) several modifications have been introduced, existing at least 10 different forms of calculating it (McGraw and Wong 1996; Nichols 1998). Since all the images were analysed by all specialists, the chosen method to use in SPSS was the two-way mixed model (all images and specialists were used), with single measurements (one measurement for each image) and absolute agreement (areas absolute value agreement). The output is a coefficient in the range [-1..1], in which the interval limits (1 or -1) are considered an absolute agreement between the measurements.

5.1.2.3 Sensitivity and Specificity analysis

As a comparative analysis for evaluating the markings matching in pixel-to-pixel basis the classical *sensitivity* (SS) and *specificity* (SP) were determined. It was assumed that in the *marked* images (figure 4.8.b) the marked pixels represent true positive results and non-marked pixels are true negative results. With this assumption, *true positive* (TP), *false positive* (FP), *true negative* (FN) and *false negative* (FN) pixels were computed in the whole analysis area and in each Wisconsin subfield (A1-A5) between the automatically *marked* images and each of the specialists marked images and also between pairs of specialists' markings. With these indicator values it is possible to compute the SS and SP using equation (5.2).

In this work the SS gives a good perception on how the markings are matching, which is not so strong when analysing the SP. Since the number non-marked pixels (TN) is usually much higher than the number of wrong marked pixels (FN), these last were masked within TN pixels. The consequence is that the SP indicator is very high and will have a low sensitivity to wrong marked pixels.

$$SS = \frac{TP}{TP + FN} \qquad SP = \frac{TN}{TN + FP} \tag{5.2}$$

5.1.2.4 Cohen's Kappa Statistics

Taking into consideration the lack of sensibility of the SP indicator for assessing underestimation of drusen, the Cohen's *kappa* Coefficient (Cohen 1960) was calculated as an alternative measurement of agreement between two raters, hereon designated by *A* and *B*. This coefficient is commonly used in these types of studies and calculated using equation (5.3).

$$k = \frac{P_r(a) - P_r(e)}{1 - P_r(e)} \tag{5.3}$$

It is based on variables $P_r(a)$ and $P_r(e)$, which are the probability of agreement among raters and the hypothetical probability of chance agreement, respectively. These two can be obtained using the following equations:

$$\begin{aligned} P_r(a) &= \frac{TP + TN}{Total\_Pixels} \\ P_r(e) &= \frac{(TP + FP_A)\cdot(TP + FP_B)}{Total\_Pixels} + \frac{(TN + FN_A)\cdot(TN + FN_B)}{Total\_Pixels} \end{aligned} \tag{5.4}$$

With the Cohen's *kappa* Coefficient, if the raters are in complete agreement then $k = 1$. On the other hand, if there is no agreement among the raters, then $k \leq 0$. Landis and Koch (1977) proposed to interpret *k* values according with the table 5.1. Although this was not universally accepted, since they supplied no evidence to support it, it has been frequently used as guidelines for analyses.

Table 5.1 – Cohen's *kappa* coefficient interpretation ((Landis and Koch 1977)).

k	**Interpretation**
< 0	No agreement
0.0 — 0.20	Slight agreement
0.21 — 0.40	Fair agreement
0.41 — 0.60	Moderate agreement
0.61 — 0.80	Substantial agreement
0.81 — 1.00	Almost perfect agreement

5.1.2.5 Weighted Matching Analysis

In the assessment techniques already presented, when there is the need to compare two or more analyses in between, for example to compare the automated method with a panel of specialists, the indicators are computed for each pair of analyses and the final results are averaged. However, this technique hides the variability which was observed among the specialists' markings. This variability increased the difficulty of getting a reference image to which the automated method could be compared to.

In order to evaluate the automated method in comparison with a panel of specialists, a novel method, which evaluates a weighed matching between one analysis and a set of analysis produced by a panel of specialists, was developed and published (Mora, Fonseca et al. 2009). Although, it can be used to compare two analyses, it was designed for studies with larger datasets with three or more specialists.

The developed method gathers the specialists' markings into one reference image, which is considered hereon the ground truth of the study. This image is a probability map, in which, each pixel's value is the percentage of specialists that have marked it (figure 5.2 (a,b,c,d)). Here the value in each pixel is obtained by equation (5.5).

$$P_{map}(x,y) = \frac{\sum_{specialists} analysis(x,y)}{count(specialists)} \tag{5.5}$$

This probability map is then the starting point for comparing one analysis to a statistically significant one. In this sense, the *true positive* pixels will be the ones with a majority of *true positive* analysis and the *true negative* pixels will be the ones where there is a majority of negative analysis. However, it was important to differentiate those pixels where there was absolute consensus and the ones where there was just a majority ($0.5 \leq P_{map} < 1$). Therefore, it was created two auxiliary probability maps for *negative* and *positive* pixels using the equation (5.6), obtaining a similar result as shown in figure 5.2 (e,f).

$$\begin{aligned} TruePositive\_P_{map}(x,y) &= \begin{cases} P_{map(x,y)} & if \quad P_{map}(x,y) \geq 0.5 \\ 0 & if \quad P_{map}(x,y) < 0.5 \end{cases} \\ TrueNegative\_P_{map}(x,y) &= \begin{cases} 1 - P_{map(x,y)} & if \quad P_{map}(x,y) < 0.5 \\ 0 & if \quad P_{map}(x,y) \geq 0.5 \end{cases} \end{aligned} \tag{5.6}$$

The matching between the analyses is measured similarly to the previously presented *sensitivity*, *specificity* and *Cohen's kappa* coefficient, although here the matching between two pixels is not a *true* or *false* result, but a probability which resulted from weighing the analysis with the probability map.

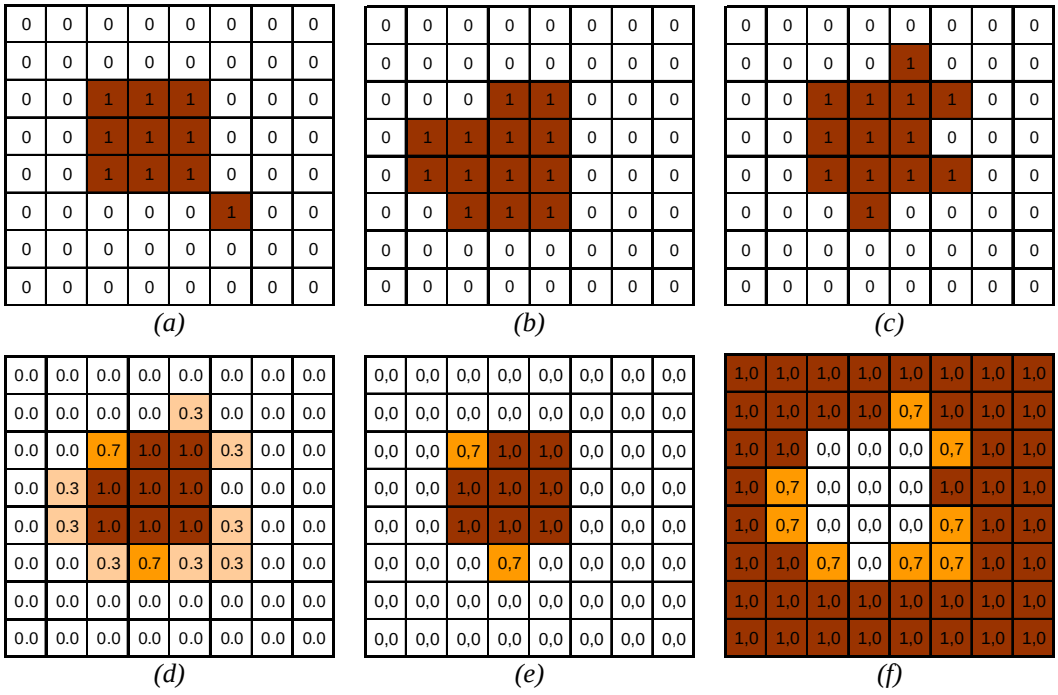


Figure 5.2 – Weighing matching analysis example. (*a*,*b*,*c*) Three gradings examples; (*d*) their probability map; (*e*) *true positi*ve probability map; (*f*) *true negati*ve probability map.

The pixel to pixel agreement is calculated with the following equations:

$$\begin{aligned} TP &= \sum_x \sum_y \left(A(x,y) * TP\_P_{map}(x,y) \right) \\ TN &= \sum_x \sum_y \left((1 - A(x,y)) * TN\_P_{map}(x,y) \right) \\ FN &= \left(\sum_x \sum_y TP\_P_{map}(x,y) \right) - TP \\ FP &= \left(\sum_x \sum_y \left(TN\_P_{map}(x,y) \right) \right) - TN \end{aligned} \tag{5.7}$$

This novel method has the advantage of not simply averaging the results and considering only statistical significance of the majority of the positive or negative classifications for calculating the *true positive* and *false positive* pixels. This improves the matching analysis by automatically removing outlier pixels and by having rewards and penalizations with different weighs based on the probabilities map values when there is no absolute agreement. Another advantage is the performance due to its ability to be easily updated by including new gradings without the need to re-compare all the examples.

With this method it is also possible to evaluate the specialists among them, by comparing each of them to the probability map. Although, to avoid its interference in the probability map, what would incorrectly improve its performance, a *leave-one-out* strategy should be used. For each of the specialists' evaluations, the probability map should be computed without his own analysis.

5.2 - Studies

Here will be presented the four studies: TRIAL I, TRIAL II, REP I and REP II. These will be described in four topics: Objective, Methods, Results and Conclusions.

5.2.1. Study I – TRIAL I

Objective:

Quantitatively compare the AD3RI drusen detection method with the specialists' markings, in order to evaluate if the results produced are comparable to the specialists.

Methods:

For this study four ophthalmologists (OP1, OP2, OP3, OP4) and four trained graders (TG1, TG2, TG3, TG4) were invited to mark a set of 22 images previously selected from the red-free and colour datasets supplied by the University of Aberdeen and from the colour dataset supplied from the ECAM study. The dataset contained images with different levels of drusen from few small drusen to large confluent drusen, and one control image which didn't contain any drusen. These specialists marked all the 22 images using a protocol (Annex B) which included the following guidelines:

- Use the MD3RI as the marking software;
- All specialists should analyze the same ROI;
- Do not overuse the illumination and contrast settings; and
- Mark the image preferably looking at its' green channel.

In figure 5.3, figure 5.4 and figure 5.5 are presented three examples of images marked by the specialists and by the AD3RI. The third one is a control that hasn't got drusen in the macular region and due to the variability observed it was considered an outlier.

To assess the variability among the specialists, the group of four trained graders was invited to repeat twice their markings in a selection of seven images.

After collecting all the specialists' analyses, a study was undertaken to identify outlier images in this dataset. The ones on which the analyses were performed by the specialists had significant differences among them, requiring a re-analysis in order to identify the causes for the dissimilarities and correct them accordingly. As this procedure was not undertaken, an outlier identification policy was implemented. The images with *coefficient of variation* higher than 50%, were considered to be outliers in this study.

The similarity between the AD3RI and the specialists' analyses was assessed using the areas comparison, *intra-class correlation coefficient*, *sensitivity* and *specificity* analysis, Cohen's *kappa* and the weighed matching analysis, which will be presented next.

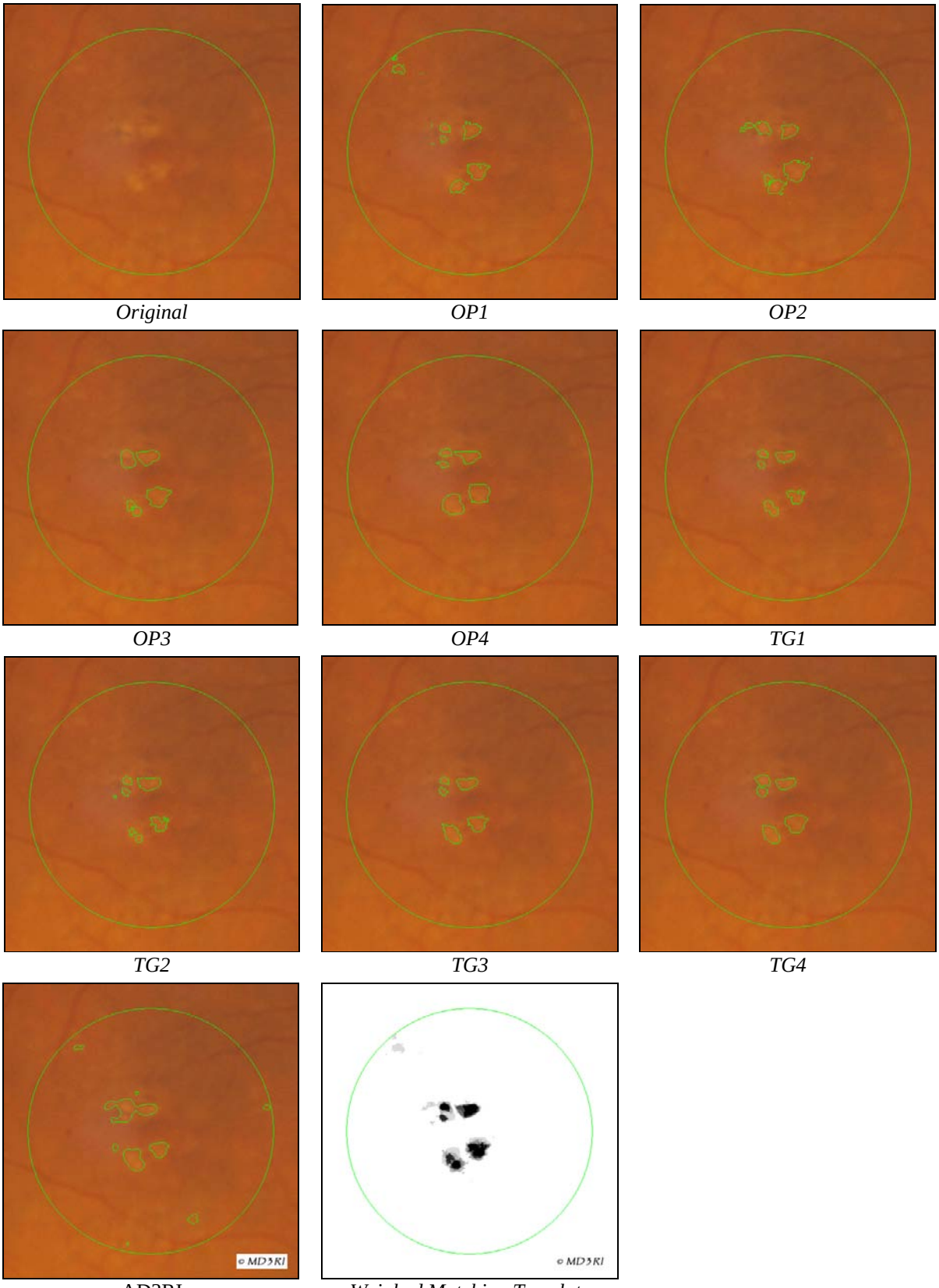


Figure 5.3 – Examples of image #1 marked by the specialists, by the AD3RI and the image template for the weighed matching analysis.

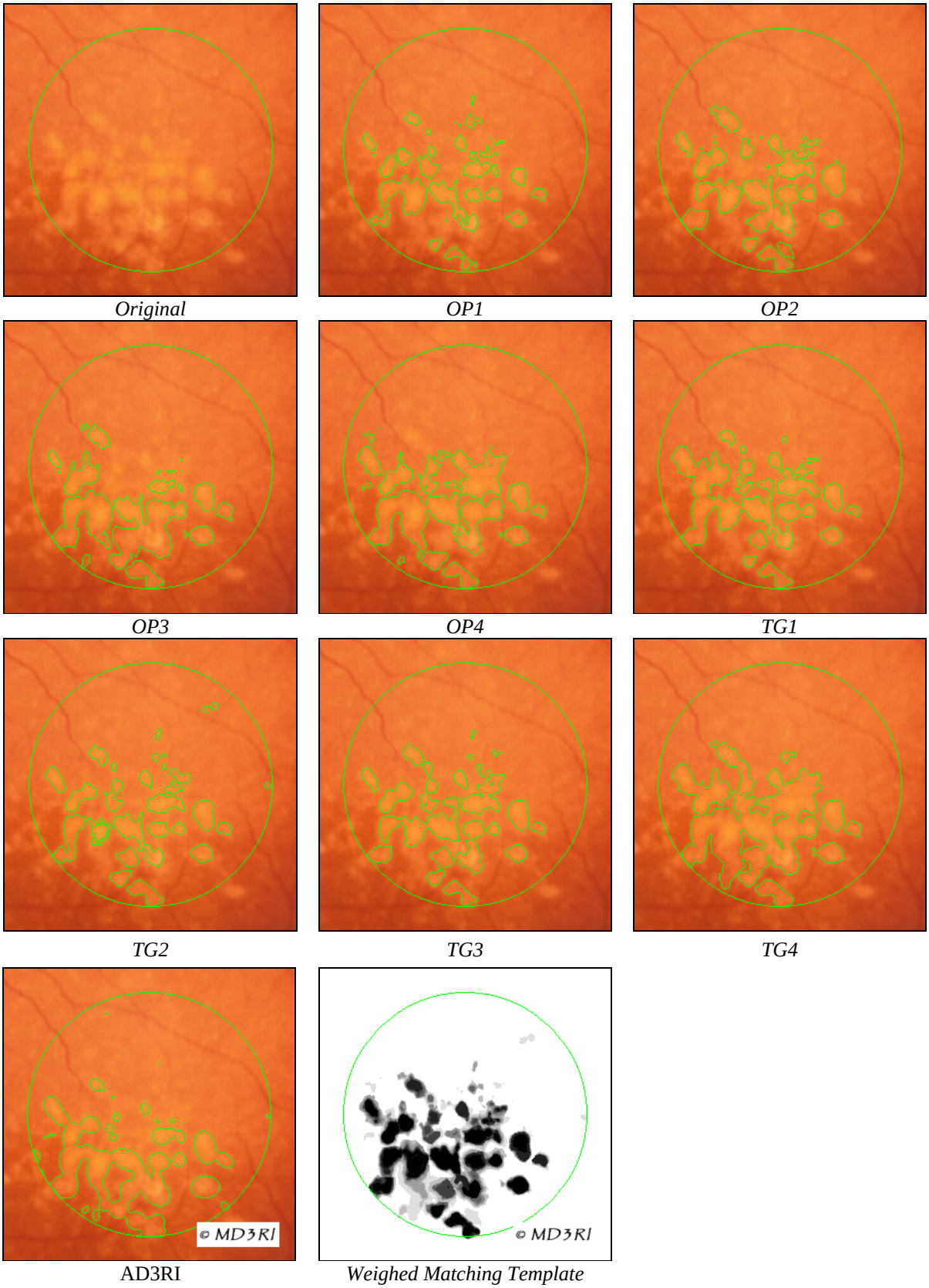


Figure 5.4 – Examples of image #20 marked by the specialists, by the AD3RI and the image template for the weighed matching analysis.

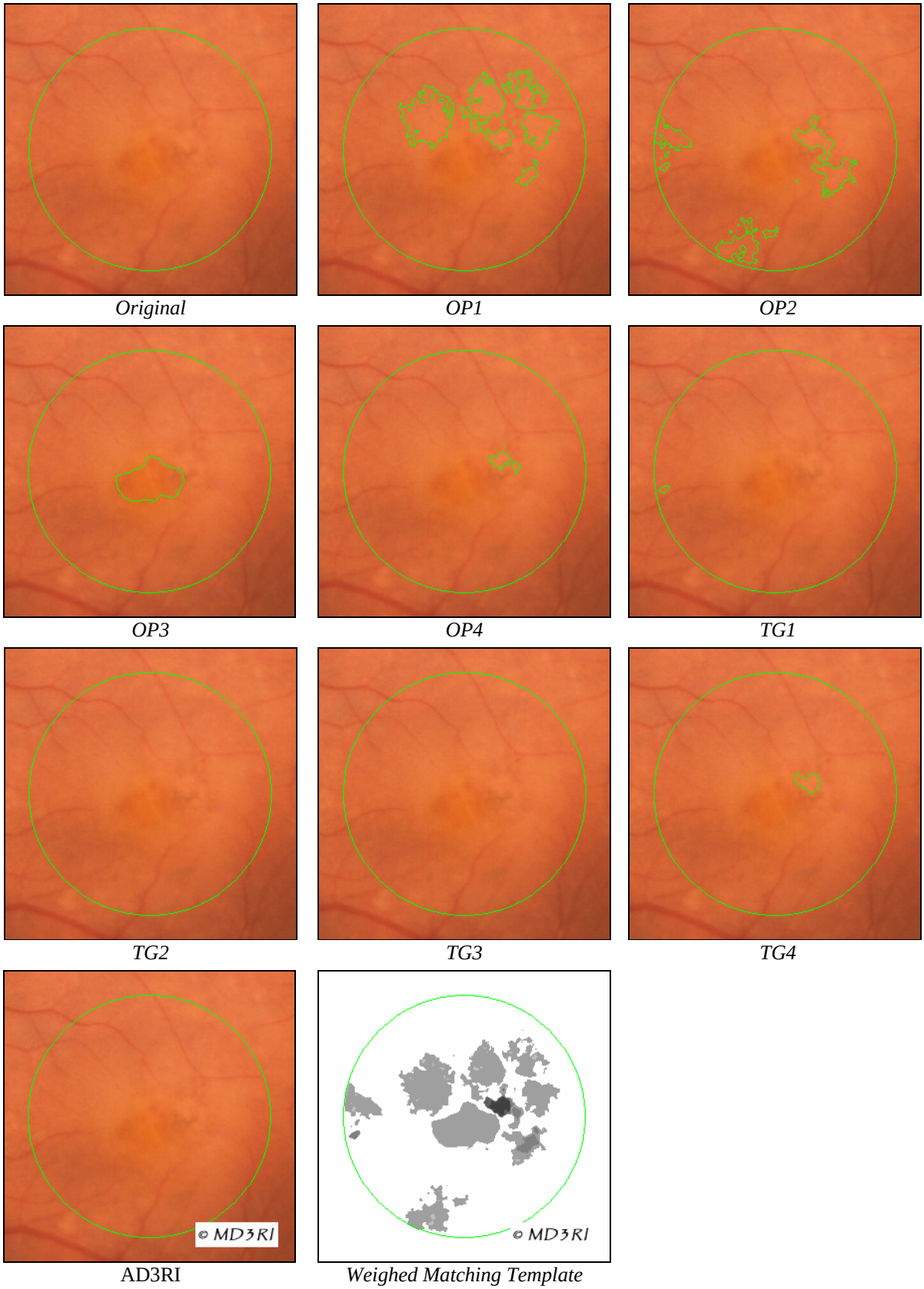
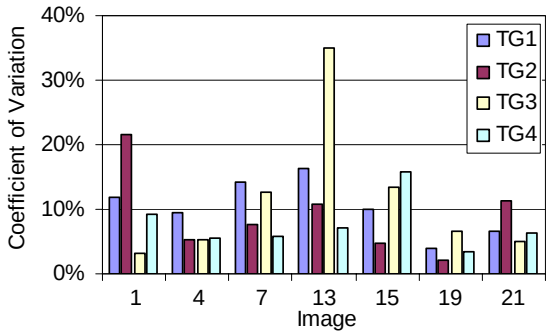


Figure 5.5 – Examples of image #22 marked by the specialists, by the AD3RI and the image template for the weighed matching analysis.

Results:

Variability among specialists:

After collecting all the three repeated analyses from the four trained graders the coefficient of variation was computed for each pair *image-specialist* and displayed on the table and plot shown in figure 5.6. The overall summary values are the mean values for the columns and rows. The ICC was also computed to evaluate the agreement between the repeated measures. The overall ICC is obtained with all the repeated images from all specialists.



	CV / Image #							Summary		
Spec	**1**	**4**	**7**	**13**	**15**	**19**	**21**	**Mean(CV)**	**Std(CV)**	**ICC**
TG1	12%	10%	14%	16%	10%	4%	7%	10%	4%	0,994
TG2	22%	5%	8%	11%	5%	2%	11%	9%	6%	0,996
TG3	3%	5%	13%	35%	13%	7%	5%	12%	11%	0,988
TG4	9%	6%	6%	7%	16%	3%	6%	8%	4%	0,997
Mean(CV)	11%	6%	10%	17%	11%	4%	7%	10%	6%	0,965
Std(CV)	8%	2%	4%	12%	5%	2%	3%			

Figure 5.6 – Variability analysis among the four trained graders for three repeated markings of each image. Plot showing the variability and table showing the data details and summary.

The mean variability observed, 10%, is in accordance with the expected value. Some of these results were influenced by the fact that some of the repetitions were made in different computers, which was the case for TG1 and TG3 who obtained the higher variabilities. The ICC showed a high agreement between the repeated measurements and all the specialists.

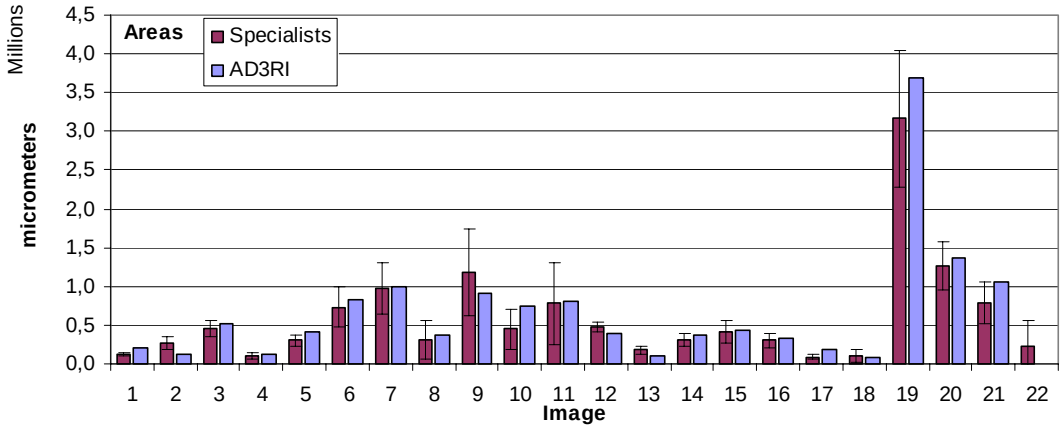
Outliers:

In this study it was considered as outliers the images where the *coefficient of variation* (CV) of the total area among the different specialists exceeded 50%. As it can be seen in figure 5.7, these were images: #8, #10, #11, #18 and #20.

Areas:

The total areas and the subfields areas defined in the Wisconsin grading system were calculated and displayed in the following figures and tables. From figure 5.7 it can be

observed that AD3RI obtained in 13 of the 17 images a total area within the maximum and minimum values observed among the specialists, knowing that 3 images contained a higher area and one the lower area.

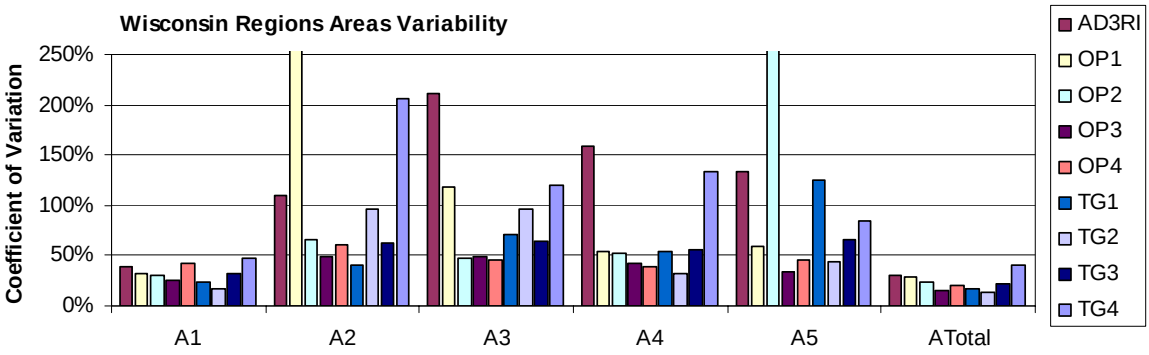


	Specialists										
#	OP1	OP2	OP3	OP4	TG1	TG2	TG3	TG4	AD3RI	Mean	CV
1	128465	155977	144091	157651	89122	94647	115462	142807	209607	128528	21%
2	225627	311732	232916	295217	201424	240092	189124	438896	126994	266879	30%
3	586222	314348	406565	548729	488080	302434	432077	500668	521333	447390	23%
4	190269	92880	60924	85159	88259	106012	106575	40691	119933	96346	46%
5	349451	260328	304151	351322	143877	245211	351913	401841	408685	301012	27%
6	756280	411010	533950	753099	718388	641208	708901	1293243	839263	727010	36%
7	918989	673398	684891	1280440	725169	1020761	868710	1613769	990815	973266	34%
8	714808	234682	6264	315092	251595	213042	152506	668282	364637	319534	77%
9	977434	974850	1028474	1425326	1064170	272227	1455660	2241263	908741	1179925	48%
10	543191	124949	299212	177363	535413	348470	717905	844544	754538	448881	57%
11	146529	273965	532911	1257297	1432789	374046	1423028	857507	806876	787259	67%
12	418019	390954	475732	427625	472149	469480	584296	511717	396520	468747	13%
13	85737	237997	160266	160484	202625	158155	260851	187777	96873	181737	30%
14	250625	251891	256839	395212	265757	319840	462643	249186	378066	306499	26%
15	101932	368082	593445	475884	404296	463153	454004	513470	432886	421783	35%
16	133588	375250	305750	340543	241748	279891	321127	415370	338395	301658	29%
17	84462	110583	76522	28192	56960	89928	134288	96084	180547	84628	38%
18	70228	78555	142012	12892	54076	97768	118636	290143	87785	108039	77%
19	2194342	1783314	3605535	3218430	3175841	2949362	3948080	4463476	3687425	3167298	28%
20	824459	1146490	1296921	1449438	1198373	1094906	1180334	1921303	1366396	1264028	25%
21	656441	1113641	641628	503774	475216	788557	1004345	1134859	1064531	789808	34%
22	880540	518527	322733	48076	9403	0	0	54294	0	229197	141%

Figure 5.7 – Graphical comparison of areas between AD3RI and the specialists (*represented by their mean, maximum and minimum values*) and total areas details for each image (*outliers in grey*).

To evaluate the variability in the areas for each specialist a *leave-one-out* strategy was used. The specialist and the AD3RI areas were left out from the mean calculation.

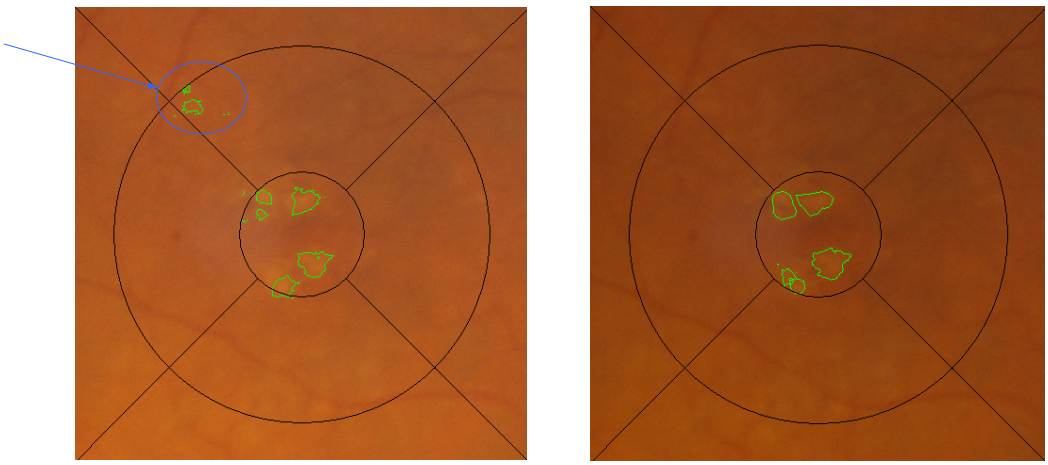
When observing the variability along the five subfields of the Wisconsin circle and the whole area it is observed a much higher discrepancy in the areas' values (figure 5.8). The AD3RI is consistently more distant from the mean value than the other specialists, only on some subfields it is more similar (A1 and ATotal). When analysing the areas details it can be identified that it is consistently generating a higher area than expected.



	A1	A2	A3	A4	A5	A1..A5	ATotal
AD3RI	39%	110%	211%	160%	134%	131%	30%
OP1	32%	11800%	118%	55%	59%	2413%	28%
OP2	30%	67%	47%	53%	498%	139%	24%
OP3	26%	49%	48%	43%	34%	40%	15%
OP4	42%	60%	46%	40%	45%	47%	20%
TG1	24%	41%	71%	54%	125%	63%	16%
TG2	17%	96%	96%	32%	45%	57%	14%
TG3	33%	62%	65%	56%	65%	56%	22%
TG4	46%	206%	120%	133%	85%	118%	41%

Figure 5.8 – Mean area variability along the Wisconsin subfields and whole area.

The overshot values for some specialists are justified in some images with low areas, as pointed out in figure 5.9. In this case and in particular in the superior subfield (A2), since OP1 had marked drusen and the others had only marked a few pixels as drusen (mean ≈ 0), his variability will tend to infinity. With the *leave-one-out* strategy some overshot markings were removed from the dataset generating different mean values which provoked this effect.



(a) *(b)*

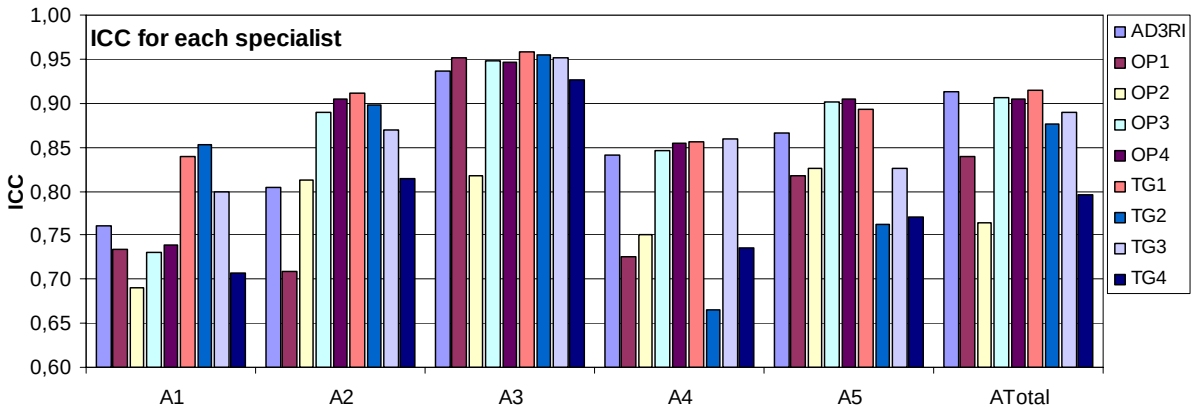
Figure 5.9 –Image #1 marked by specialists OP1 and OP3.

Intra-Class Correlation Coefficient:

The ICC was evaluated in paired comparison between AD3RI and each of the specialists. They were also evaluated among each other in order to assess their score when compared to the others and to establish a reference score to which the AD3RI could be compared to. The mean ICC values were then computed and presented in figure 5.10.

The agreement between the areas evaluated using the ICC showed a higher similarity of AD3RI in comparison to the other specialists. It also showed that the specialists among them are in some cases more disparate than the AD3RI compared to the others. The specialist *OP2* showed consistently lower agreement with the others. The AD3RI in the total area presented a similar ICC value as did TG1, which obtained the higher ICC score.

In the previous area analysis, it was presented a high variability of AD3RI in relation to the other specialists. That analysis is a comparison among the mean area values, whose outlier images are having direct influence on the results. Here, that effect is naturally solved by the ICC calculation generating much more reliable results.



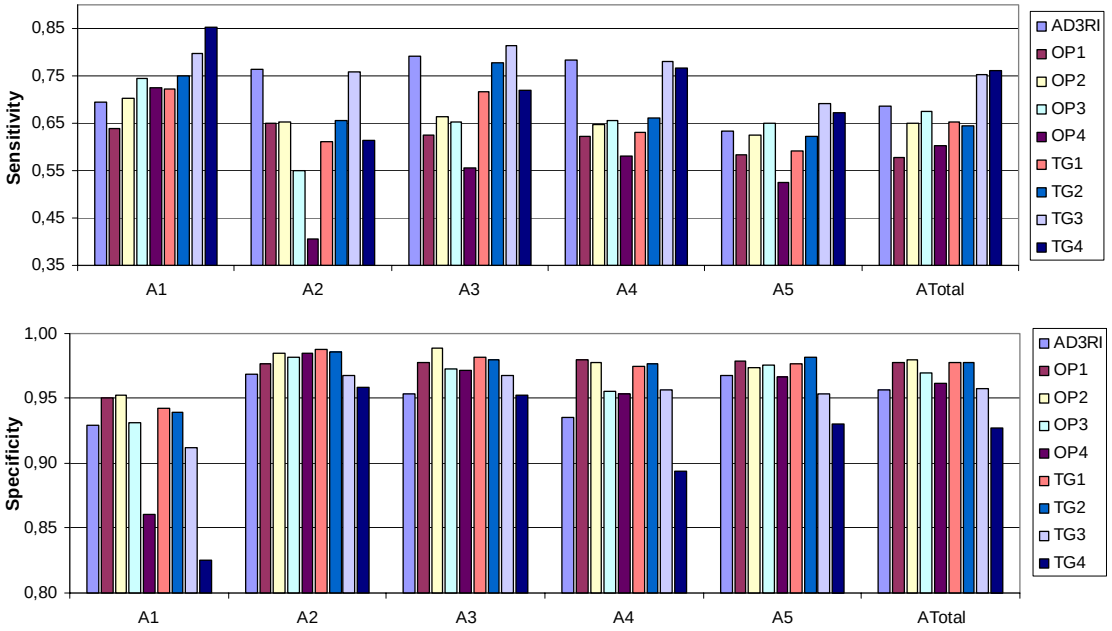
	A1	A2	A3	A4	A5	A1..A5	ATotal
AD3RI	0,761	0,804	0,936	0,840	0,866	0,841	0,913
OP1	0,734	0,709	0,952	0,725	0,818	0,788	0,840
OP2	0,691	0,813	0,817	0,751	0,826	0,780	0,765
OP3	0,730	0,889	0,948	0,846	0,902	0,863	0,906
OP4	0,739	0,904	0,947	0,854	0,905	0,870	0,905
TG1	0,839	0,912	0,958	0,855	0,892	0,891	0,914
TG2	0,852	0,898	0,954	0,666	0,763	0,827	0,875
TG3	0,800	0,869	0,952	0,859	0,825	0,861	0,889
TG4	0,707	0,815	0,927	0,736	0,770	0,791	0,796

Figure 5.10 – Mean ICC values for the dataset without outliers in the Wisconsin subfields and the total area.

Sensitivity and specificity Analysis:

The *true positive*, *false positive*, *true negative* and *false negative* pixels were computed between pairs of analysis and were separated in the five Wisconsin subfields and total area. The *sensitivity* (SS) and *specificity* (SP) were then calculated for each pair of analysis; the mean value for each pair of specialists was calculated, and finally for each specialist it was calculated the mean score with each of the other specialists. The outliers were intentionally left out from this analysis. The results are presented in figure 5.11.

Here, it is important not only the areas, but also if the pixels were marked in the correct location. Analysing the SS, it can be observed that the AD3RI obtained a good performance compared to the other specialists. The SS was even the highest in A2 subfield and was consistent within the values exhibited by the specialists among them. The SP performance was moderate, benefitting from the variability within the specialists. Among the specialists the outlier was OP4, especially in the subfield areas.

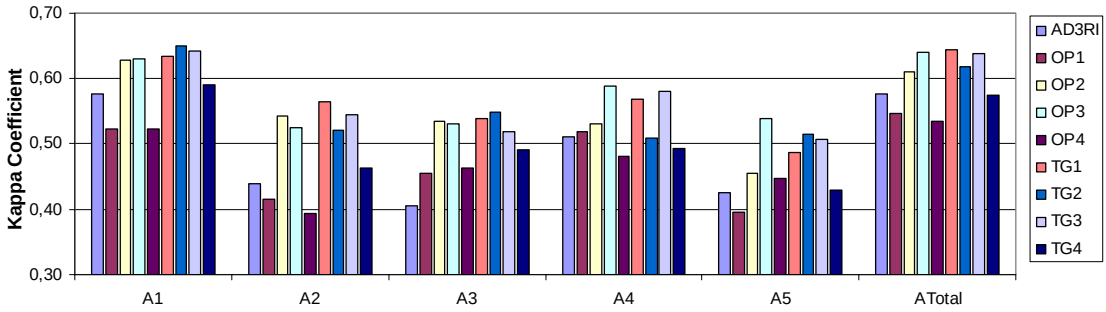


	Sensitivity							Specificity						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**	**A1**	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
AD3RI	0,694	0,764	0,793	0,783	0,634	0,734	0,686	0,929	0,969	0,953	0,936	0,968	0,951	0,957
OP1	0,640	0,651	0,625	0,624	0,583	0,624	0,577	0,951	0,977	0,978	0,980	0,979	0,973	0,977
OP2	0,703	0,652	0,663	0,647	0,625	0,658	0,649	0,952	0,985	0,989	0,978	0,974	0,976	0,979
OP3	0,744	0,549	0,652	0,657	0,649	0,650	0,675	0,931	0,982	0,973	0,955	0,975	0,963	0,969
OP4	0,724	0,404	0,556	0,582	0,524	0,558	0,604	0,861	0,985	0,972	0,953	0,967	0,948	0,962
TG1	0,722	0,611	0,716	0,630	0,591	0,654	0,653	0,942	0,987	0,981	0,974	0,977	0,972	0,977
TG2	0,751	0,654	0,779	0,661	0,621	0,693	0,645	0,940	0,986	0,980	0,977	0,982	0,973	0,978
TG3	0,797	0,760	0,813	0,779	0,691	0,768	0,752	0,913	0,968	0,968	0,957	0,953	0,952	0,958
TG4	0,852	0,615	0,719	0,766	0,672	0,725	0,762	0,825	0,959	0,953	0,894	0,931	0,912	0,927

Figure 5.11 –Mean Sensitivity and Specificity Analysis in the Wisconsin subfields and total area.

Cohen's Kappa Statistics:

The *kappa* coefficient calculus was similar to the previous analysis. It was performed for each pair of analysis and its mean value was calculated for each specialist. The performance measured by the *kappa* coefficient showed to be moderate, although within the values obtained by the specialists. Analysing these values using the *kappa* interpretation shown in table 5.1, the AD3RI showed a *moderate agreement* with the specialists ($0.4 \leq k < 0.6$) and some of the specialists showed a *substantial agreement* with the others ($0.6 \leq k < 0.8$).



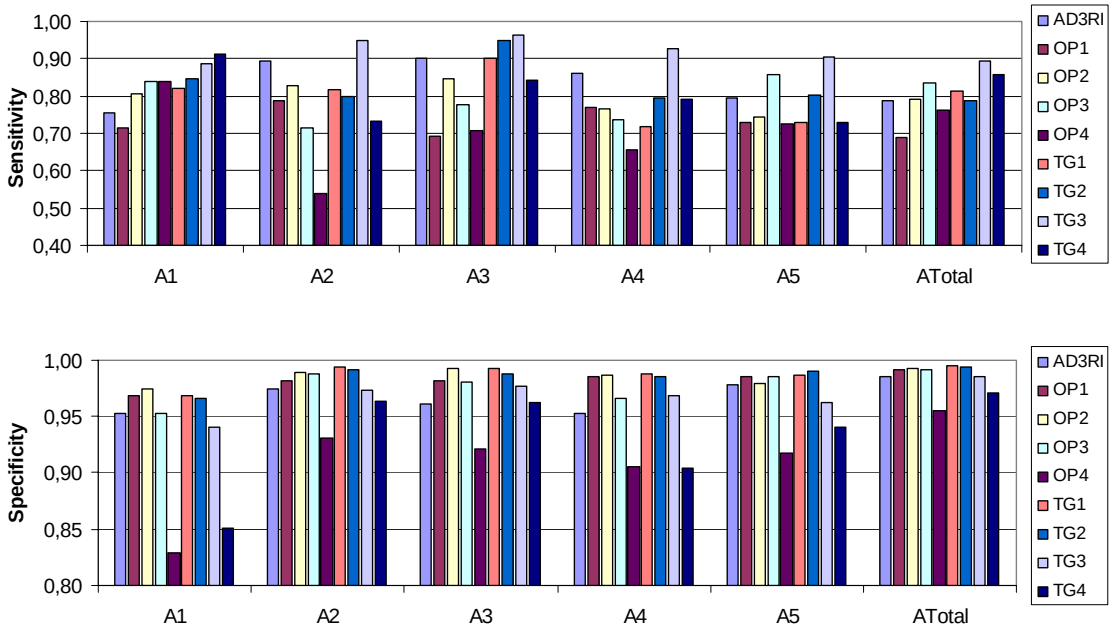
	Kappa Coefficient						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
AD3RI	0,577	0,440	0,405	0,511	0,426	0,472	0,577
OP1	0,523	0,415	0,455	0,518	0,395	0,461	0,547
OP2	0,628	0,542	0,535	0,530	0,456	0,538	0,611
OP3	0,630	0,525	0,531	0,588	0,540	0,563	0,640
OP4	0,524	0,394	0,464	0,481	0,448	0,462	0,535
TG1	0,634	0,565	0,538	0,570	0,488	0,559	0,644
TG2	0,651	0,520	0,549	0,510	0,515	0,549	0,618
TG3	0,642	0,546	0,519	0,581	0,508	0,559	0,637
TG4	0,590	0,463	0,492	0,494	0,430	0,494	0,575

Figure 5.12 – Kappa Coefficient in the Wisconsin subfields and total area.

Weighed Matching Analysis:

The probabilities maps were computed for each specialist using a *leave-one-out* strategy to avoid the influence of the current specialist in the template generation. The *true positive*, *false positive*, *true negative* and *false negative* pixels were computed between each analysis and the current template and were separated in the five Wisconsin subfields and total area. For each image it was analyzed the *sensitivity*, *specificity* and *kappa* coefficient.

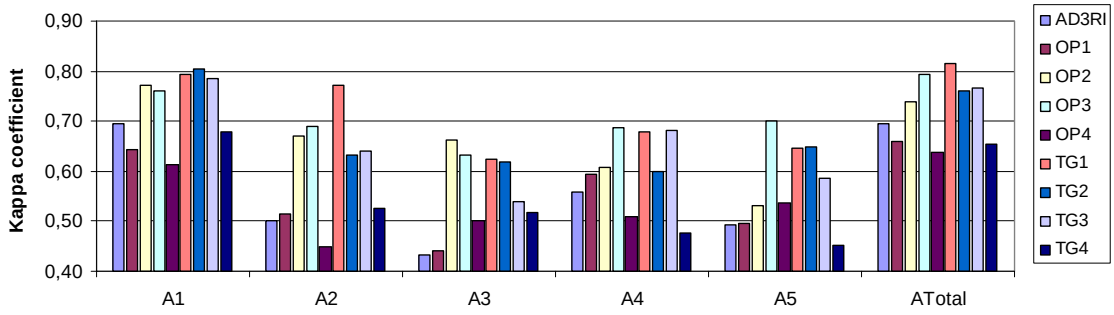
From the values obtained for the SS and SP it is not evidenced any significant change in the AD3RI performance, what is not true for some other specialists (for example, TG4 decreased its performance in comparison with its pairs). However, the AD3RI obtained in all the subfields a performance similar to the ones exhibited by the specialists, as it was inside the limits found for the specialists.



	Sensitivity							Specificity						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**	**A1**	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
AD3RI	0,754	0,893	0,903	0,860	0,793	0,841	0,787	0,953	0,975	0,961	0,952	0,978	0,964	0,985
OP1	0,713	0,786	0,691	0,770	0,730	0,738	0,690	0,968	0,981	0,982	0,986	0,986	0,980	0,992
OP2	0,806	0,827	0,848	0,767	0,744	0,798	0,793	0,974	0,989	0,993	0,987	0,980	0,985	0,993
OP3	0,840	0,716	0,775	0,736	0,857	0,785	0,834	0,953	0,988	0,981	0,966	0,986	0,975	0,991
OP4	0,840	0,540	0,708	0,657	0,725	0,694	0,762	0,829	0,930	0,922	0,905	0,918	0,901	0,955
TG1	0,822	0,818	0,901	0,718	0,729	0,797	0,812	0,969	0,994	0,993	0,988	0,986	0,986	0,995
TG2	0,845	0,797	0,948	0,797	0,804	0,838	0,787	0,967	0,992	0,987	0,986	0,990	0,984	0,994
TG3	0,887	0,949	0,964	0,926	0,904	0,926	0,895	0,940	0,974	0,977	0,969	0,962	0,964	0,986
TG4	0,913	0,732	0,841	0,790	0,729	0,801	0,856	0,851	0,964	0,962	0,904	0,940	0,924	0,971

Figure 5.13 –Mean Sensitivity and Specificity analysis in the Wisconsin subfields and total area, using the weight matching analysis.

The *kappa* coefficient calculated using the Weighed Matching Analysis showed also the similarities between the AD3RI and the specialists' markings. As the template contains partially true and false pixels which are also counted the overall k-value, this latter has increased, making the AD3RI to exhibit a *substantial agreement* with the template.



	Kappa						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
AD3RI	0,70	0,50	0,43	0,56	0,49	0,54	0,70
OP1	0,64	0,52	0,44	0,59	0,50	0,54	0,66
OP2	0,77	0,67	0,66	0,61	0,53	0,65	0,74
OP3	0,76	0,69	0,63	0,69	0,70	0,69	0,79
OP4	0,61	0,45	0,50	0,51	0,54	0,52	0,64
TG1	0,79	0,77	0,62	0,68	0,65	0,70	0,82
TG2	0,81	0,63	0,62	0,60	0,65	0,66	0,76
TG3	0,78	0,64	0,54	0,68	0,59	0,65	0,76
TG4	0,68	0,53	0,52	0,48	0,45	0,53	0,65

Figure 5.14 – Mean Kappa Coefficient in the Wisconsin subfields and total area using the weight matching analysis.

Conclusions:

This study evaluated the agreement between the analyses performed by AD3RI and by a panel of specialists. It was also assessed the agreement among all specialists as a benchmark for the AD3RI, and the analysis repeatability for some specialists. As was expected, the subjectivity of the analysis when performed by the specialists resulted in a non-negligible variability among all the specialists, from 14% to 41% and an acceptable repeatability, between 8% and 12%. These conclusions were also confirmed with the ICC, the agreement among all the specialists ranged from 0.76 to 0.90 and the repeatability was consistently near 0.99. They can repeat their analysis without compromising the variability, but among them there are considerable differences in the analysis criteria.

The assessment of AD3RI at the area level showed that it produced similar results to the ones produced by the specialists. The variability however is higher than the majority, mainly due to a higher total area, as consequence of a more detailed analysis. The ICC obtained 0.91 is equal to the maximum among the specialists.

At pixel level the *sensitivity*, *specificity* and *kappa* coefficient obtained by the AD3RI were within the values obtained among the specialists, what was also verified using the Weighed Matching Analysis.

There were no significant differences among the analysis of the Wisconsin subfields, showing that the analysis criterion is uniform all over the image.

5.2.2. Study II – TRIAL II

Objective:

To quantitatively compare the AD3RI software with other specialists using a different dataset.

Methods:

For this study one ophthalmologist (OP1) and one trained grader (TG1) were invited to mark a set of 39 images previously selected from the colour datasets supplied by the University of Aberdeen, and from the colour dataset supplied from the ECAM study. As in the previous dataset it was tried to have diversity of the types of drusen present in the dataset.

The protocol for analysing the images was similar to the previous one, except that the variability intra-rater wasn't evaluated.

In this study the similarity between the automated drusen detection and the specialists analyses was assessed using the areas comparison, *intra-class correlation coefficient*, *sensitivity* and *specificity* analysis and Cohen's *kappa* statistics, which all will be presented next. As there are only two specialists involved in this study, the weighted matching analysis is not applicable. This method is more suitable to be used in studies where the comparison is with three or more specialists.

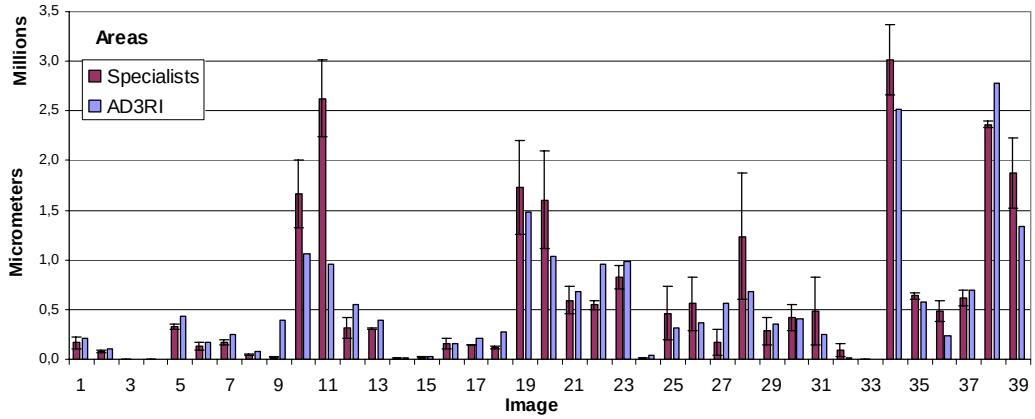
Results:

Outliers:

In this study it was considered as outliers the images whose *coefficient of variation* of the total area among the different specialists (mean value) exceeded 50%, as well as some additional images which contained severe macular changes not related with drusen (macular hole and geographic atrophy) (#8, #10, #11). As it can be seen in figure 5.15, these were the images: #8, #10, #11, #24, #25, #26, #27, #28, #29, #31 and #32.

Areas:

The total area and the five subfield areas were calculated and displayed in following figures and tables. From figure 5.15 it can be observed that AD3RI obtained in 10 of the 30 images a total area within the maximum and minimum values observed among the specialists, 13 images containing higher areas and five lower areas.



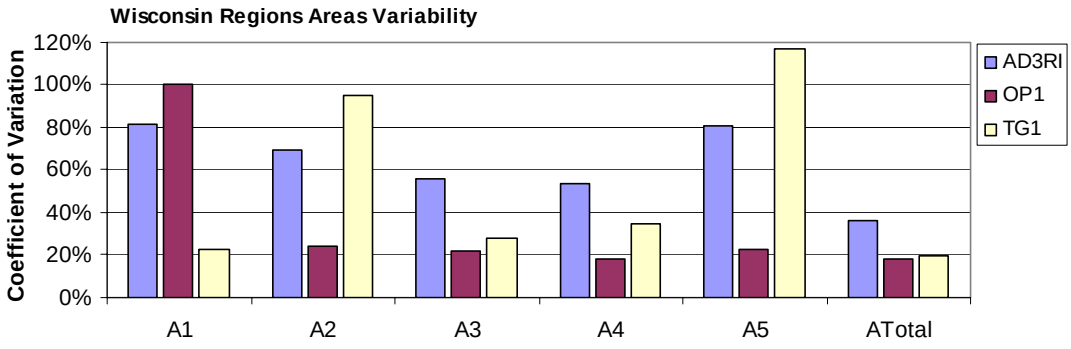
	Specialists						Automated	
#	OP1	TG1	Mean	CV(OP1)	CV(TG1)	Mean(CV)	AD3RI	CV
1	222107	107315	164711	53%	26%	40%	212453	29%
2	64920	88493	76706	13%	18%	16%	110873	45%
3	0	0	0	0%	0%	0%	0	0%
4	0	0	0	0%	0%	0%	0	0%
5	307693	355761	331727	7%	8%	7%	432394	30%
6	86568	176574	131571	25%	52%	39%	164738	25%
7	146816	199512	173164	13%	18%	16%	254293	47%
8	32975	52772	42873	19%	30%	24%	78190	82%
9	15950	20957	18453	12%	16%	14%	389100	2009%
10	1319177	2005089	1662133	17%	26%	22%	1063560	36%
11	2242407	3011505	2626956	13%	17%	15%	956392	64%
12	419982	206791	313386	52%	25%	38%	550766	76%
13	310401	295573	302987	3%	2%	2%	386743	28%
14	11985	5732	8858	55%	26%	40%	15893	79%
15	19735	11429	15582	36%	21%	29%	29665	90%
16	103541	208341	155941	25%	51%	38%	162590	4%
17	143033	143222	143128	0%	0%	0%	204576	43%
18	135771	101048	118409	17%	13%	15%	278724	135%
19	1262038	2207550	1734794	21%	37%	29%	1478745	15%
20	1109077	2098390	1603734	24%	45%	34%	1035198	35%
21	730905	453577	592241	31%	19%	25%	681599	15%
22	588351	499986	544168	9%	8%	8%	950558	75%
23	713056	948412	830734	12%	17%	14%	979604	18%
24	0	14807	7404	50%	100%	75%	34597	367%
25	194062	730571	462316	37%	138%	87%	320214	31%
26	292923	826800	559862	32%	91%	62%	362140	35%
27	33746	298587	166166	44%	392%	218%	559042	236%
28	604301	1869774	1237037	34%	105%	69%	675890	45%
29	146779	425778	286279	33%	95%	64%	357124	25%
30	284872	547418	416145	24%	46%	35%	407201	2%
31	148267	828256	488262	41%	229%	135%	246024	50%
32	24212	156066	90139	42%	272%	157%	13421	85%
33	0	0	0	0%	0%	0%	0	0%
34	2664847	3365627	3015237	10%	13%	12%	2513389	17%
35	609196	671345	640270	5%	5%	5%	581707	9%
36	382056	591348	486702	18%	27%	23%	241431	50%
37	535572	697149	616360	12%	15%	13%	696384	13%
38	2394331	2330212	2362272	1%	1%	1%	2781070	18%
39	1521675	2230960	1876318	16%	23%	20%	1338069	29%

Figure 5.15 – Graphical comparison of areas between AD3RI and the specialists (*represented by their mean, maximum and minimum values*) and total areas details for each image (*outliers in grey*).

To compute the *coefficient of variation*, each specialist results were compared to the others by computing their differences and by dividing them by the other specialists' results. The final *Mean(CV)* is the mean between the two specialists *coefficient of variation*.

When observing the variability along the five subfields of the Wisconsin circle and the whole area it is observed a higher discrepancy in the area values (figure 5.16). The AD3RI is consistently more distant from the mean value than the other specialists. However, in some subfields it is within the specialists values (A1 and A2). When analysing the areas details it can be identified that it is consistently generating a higher area than expected.

The overshot values for some subfields are justified by images where the specialists have marked very few drusen and the AD3RI detected a higher number. In these cases the proportion between them will be high and consequently generate higher variabilities.



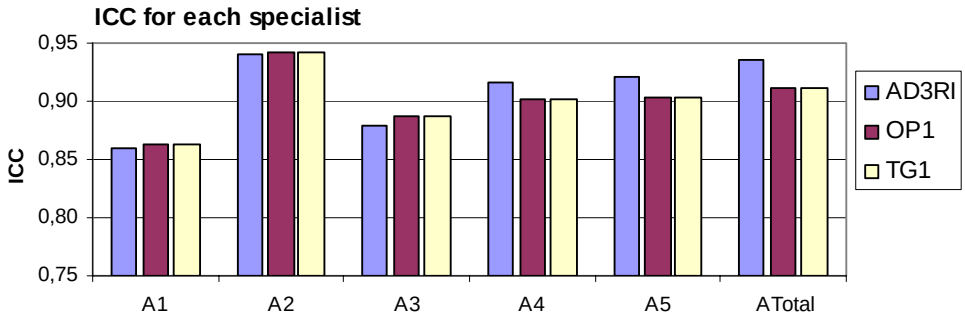
	A1	A2	A3	A4	A5	A1..A5	ATotal
AD3RI	82%	69%	56%	54%	81%	68%	36%
OP1	101%	24%	22%	18%	22%	37%	18%
TG1	22%	95%	28%	34%	117%	59%	20%

Figure 5.16 – Mean area variability along the Wisconsin subfields and whole area.

Intra-Class Correlation Coefficient:

The agreement between the area values evaluated by the ICC showed a higher similarity between AD3RI and the other two specialists. The ICC values for the AD3RI are the mean ICC evaluated with each of the specialists independently. The scores of the specialists are equal because the ICC evaluates the similarity of pairs of measurements despite their order.

Analysing the results in figure 5.17 it can be noticed that in subfields A4, A5 and ATotal the score obtained by AD3RI was even higher then the specialists between them.



	A1	A2	A3	A4	A5	A1..A5	ATotal
AD3RI	0,860	0,940	0,880	0,916	0,920	0,903	0,935
OP1	0,863	0,942	0,888	0,902	0,903	0,900	0,912
TG1	0,863	0,942	0,888	0,902	0,903	0,900	0,912

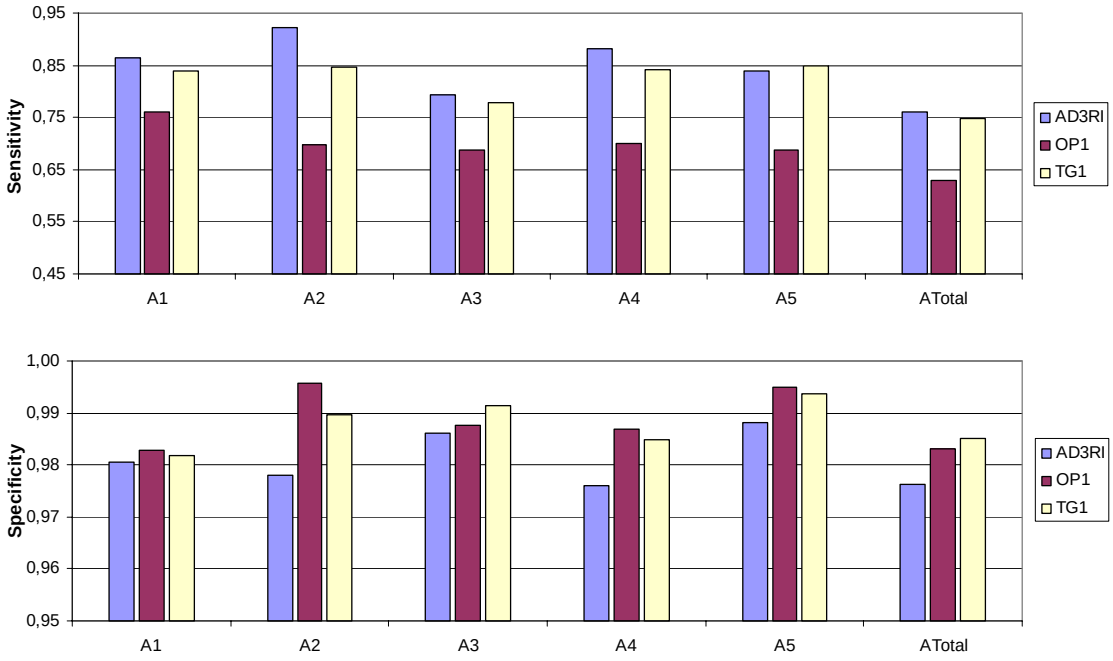
Figure 5.17 – Mean ICC values for the dataset without outliers in the Wisconsin subfields and the total area.

Sensitivity and specificity Analysis:

The *true positive*, *false positive*, *true negative* and *false negative* pixels were computed using the same procedure as in the previous study. The *sensitivity* (SS) and *specificity* (SP) were then calculated for each pair of analyses; the values for the two specialists are evaluated only between them. The outliers were intentionally left out from this analysis. The results are presented in figure 5.18.

Analysing the SS, it can be observed that the AD3RI obtained a good performance compared to the other specialists. The SS was only below the specialists' values in subfield A5, and the mean value was above 0.7, which was considered a good agreement for these studies.

Since the AD3RI was consistently marking higher areas than the others, the SP performance was compromised, although it was within the values obtained by the specialists. As mentioned before, this indicator is influenced by the high number of *true negatives* and only a high value of *false negatives* will produce visible changes.

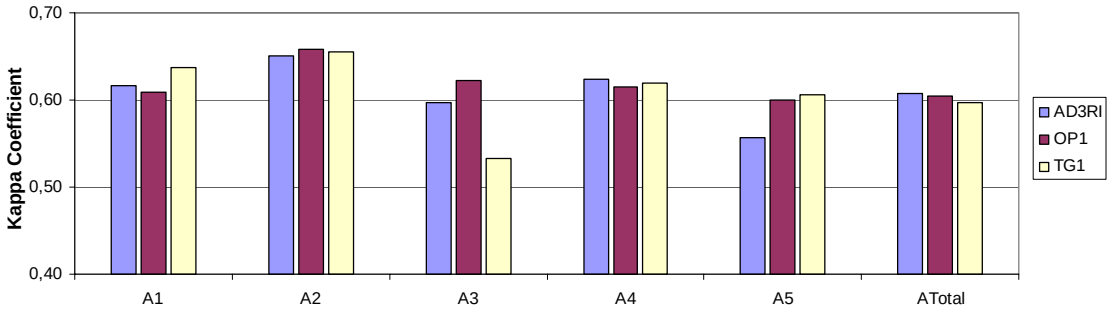


	Sensitivity							Specificity						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**	**A1**	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
AD3RI	0,864	0,921	0,794	0,881	0,839	0,860	0,761	0,981	0,978	0,986	0,976	0,988	0,982	0,976
OP1	0,759	0,698	0,687	0,701	0,687	0,707	0,629	0,983	0,996	0,988	0,987	0,995	0,990	0,983
TG1	0,838	0,848	0,779	0,843	0,849	0,831	0,748	0,982	0,990	0,991	0,985	0,994	0,988	0,985

Figure 5.18 – Analysis of mean *Sensitivity* and *Specificity* in the Wisconsin subfields and total area.

Cohen's Kappa Statistics:

The Cohen's *kappa* coefficient was calculated using the same procedure as in the previous study, and obtained the results shown in figure 5.19. The overall agreement was even higher than the previous study, as all the specialists and AD3RI obtained values near the 0.6, which is the threshold between *moderate* and *substantial* agreement. The AD3RI obtained a good score being very near of obtaining a *substantial* agreement with the specialists' analyses.



	Kappa Coefficient						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
AD3RI	0,616	0,651	0,596	0,624	0,556	0,609	0,608
OP1	0,609	0,658	0,623	0,616	0,601	0,621	0,605
TG1	0,637	0,655	0,532	0,620	0,606	0,610	0,597

Figure 5.19 – Mean *Kappa* Coefficient in the Wisconsin subfields and total area.

Conclusions:

This study was undertaken to reinforce the results of the previous study. From the results presented here and besides those regions where the areas of some images were proportionally very different, which should not be considered an error, the AD3RI can be considered as capable to produce results similar to a panel of specialists. It should be noticed that in most of the important indicators such as ICC, SS and *Kappa* coefficient the AD3RI produced consistent values and was always within the values obtained by the specialists when evaluated among them.

5.2.3. Study III – REP I

Objective:

Evaluate the repeatability of the automated detection method in analysing sequences of images from the same eye.

Methods:

For this study, only the AD3RI marked the selected images. The image dataset used was the one from University of Abeerdeen (24 images) since it contained two or more images of the same eye acquired in the same conditions.

From this dataset there was only one patient rejected in a total of 8 patients containing at least two images from the same eye. Eliminating these images, which didn't comply with the requisites, it was compiled a dataset containing 9 pairs of retinal images.

The procedure in this study was to analyse automatically each pair of images and evaluate the agreement between the two images (figure 5.20). After the drusen detection the agreement was assessed using the areas comparison, *intra-class correlation coefficient*, *sensitivity* and *specificity* analysis and Cohen's *kappa* statistics, which all will be presented next.

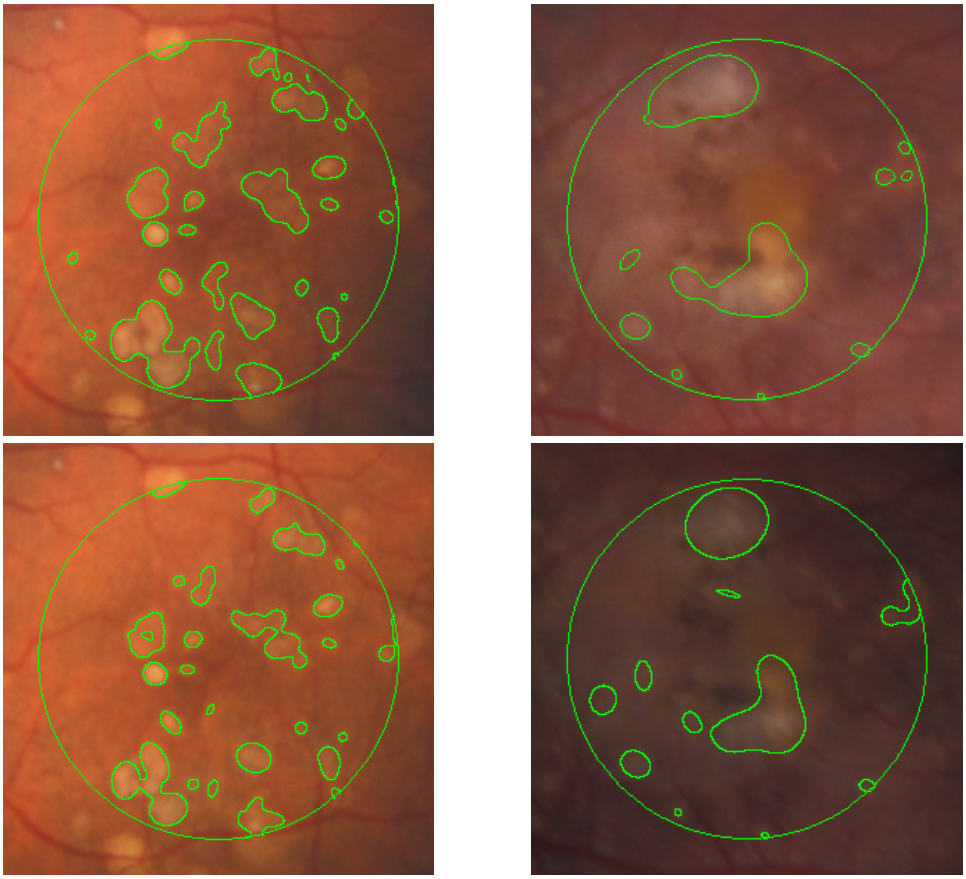
Results:

Outliers:

In this study it wasn't specified any outlier definition protocol, since the images which didn't have the required acquisition quality were previously removed from the dataset. In this sense none of the images should be classified as an outlier.

Areas:

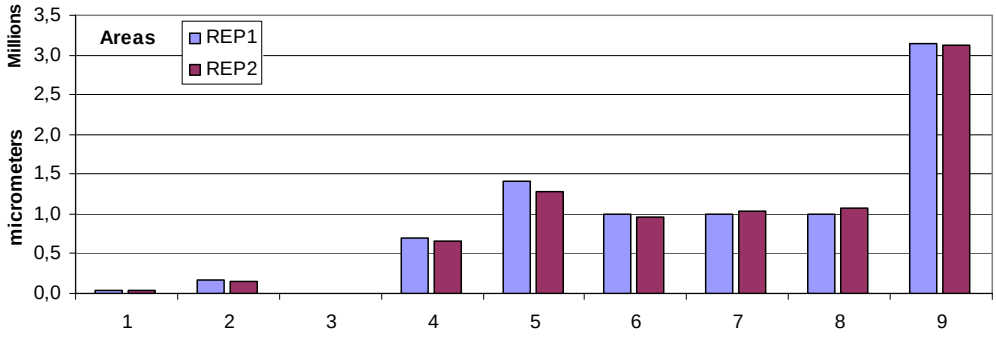
The total area and the subfields areas defined in the Wisconsin grading system were calculated and displayed in the following figures and tables. It can be observed that the number of total areas which are higher for REP1 and for REP2 are even, i.e., there is no identifiable predomination of one of the repetitions.



#5 *#6*

Figure 5.20 – Examples of repeated images and their correspondent analysis.

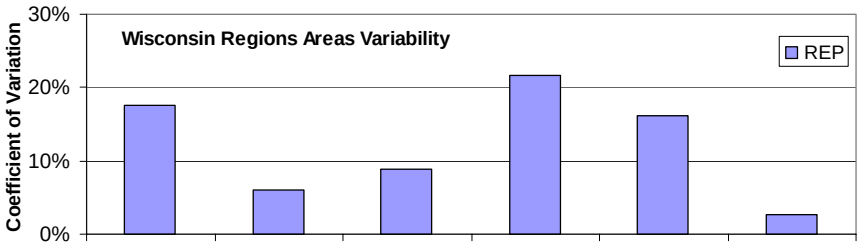
When observing the variability in the whole area and along the five subfields of the Wisconsin circle it is observed a higher discrepancy in the area values (Figure 5.21). The *coefficient of variability* in the whole area was always below 4.5% with an average of 3%,



#	REP1	REP2	Mean	CV
1	34882	32304	33593	3,8%
2	168495	157878	163187	3,3%
3	0	0	0	0,0%
4	701933	650957	676445	3,8%
5	1405229	1285791	1345510	4,4%
6	991252	958038	974645	1,7%
7	991252	1032048	1011650	2,0%
8	991252	1073452	1032352	4,0%
9	3151198	3126477	3138838	0,4%

Figure 5.21 – Graphical comparison of areas between the analysis 1 and 2 and total areas details for each image.

what can be considered as a good repeatability. While within the Wisconsin subfields it was higher, mainly due to some images which have small affected areas in these subfields. In these cases any difference in the areas' values generates a high value in the *coefficient of variability*.

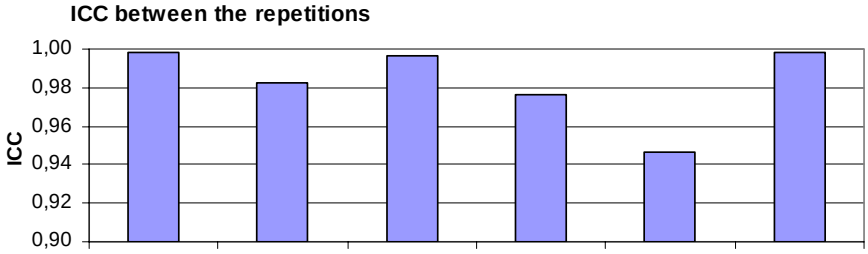


	A1	A2	A3	A4	A5	A1..A5	ATotal
REP	18%	6%	9%	22%	16%	14%	3%

Figure 5.22 – Mean area variability along the Wisconsin subfields and whole area.

Intra-Class Correlation Coefficient:

The agreement between the area values evaluated by the ICC showed a high agreement between the repetitions in comparison to the previous studies (figure 5.23). It can be observed that the mean ICC was above 0.98 which is considered a good agreement.



	A1	A2	A3	A4	A5	A1..A5	ATotal
REP	0,999	0,982	0,996	0,976	0,946	0,980	0,998

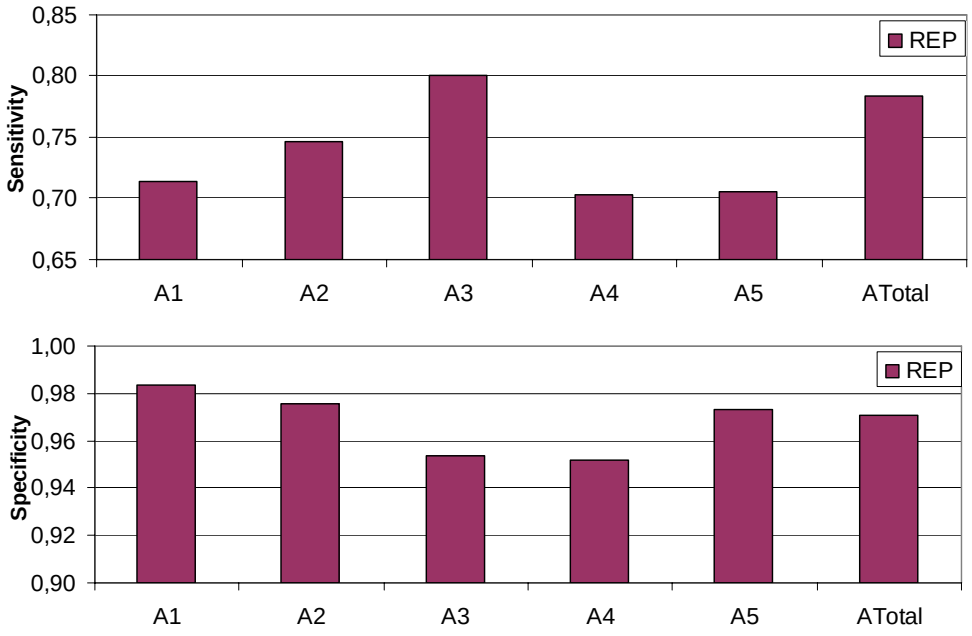
Figure 5.23 – Mean ICC values in the Wisconsin subfields and the total area.

Sensitivity and specificity Analysis:

To compute the *true positive*, *false positive*, *true negative* and *false negative* pixels it was assumed that the gold standard was the first image and therefore the repetition would be compared against it. The *sensitivity* (*SS*) and *specificity* (*SP*) were then calculated and presented in figure 5.24.

The SS achieved in the Wisconsin subfields a mean value above 0.7 and the SP was consistent above 0.95, which are considered both good results. On A3 the high SS is not associated with a high SP, which shows that the analysis criterion is similar on all the

subfields. On the total area the 0.78 and 0.97 for SS and SP, respectively, show a good repeatability of the AD3RI.

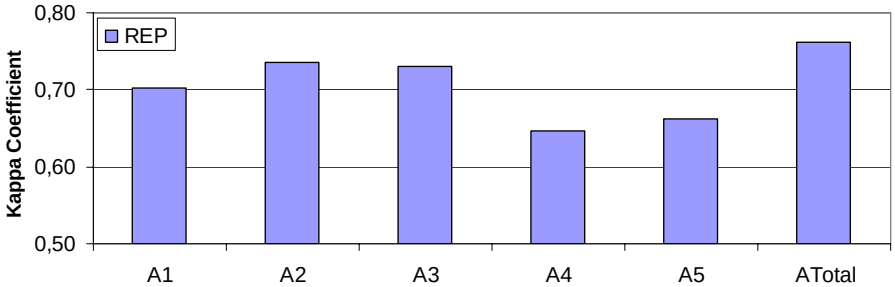


	Sensitivity							Specificity						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**	**A1**	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
REP	0,714	0,747	0,801	0,703	0,705	0,734	0,784	0,983	0,976	0,954	0,952	0,973	0,967	0,971

Figure 5.24 – *Sensitivity* and *Specificity* analysis in the Wisconsin subfields and total area.

Cohen's Kappa Statistics:

The Cohen's *kappa* coefficient was calculated using the same procedure as in the previous studies and obtained the results shown in figure 5.25. The overall agreement between repetitions was *substantial*, in all subfields. In the total area the agreement obtained the maximum value with 0.76, while the performance among the Wisconsin subfields was similar but lower.



	Kappa Coefficient						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
REP	0,702	0,736	0,730	0,647	0,661	0,695	0,761

Figure 5.25 – *Kappa* Coefficient in the Wisconsin subfields and total area.

Conclusions:

This study assessed the repeatability and the mean error of the automatic quantification drusen algorithm. The mean error was approximately 3% when analysing the whole macular area and about 14% when analysing the Wisconsin subfields. The ICC obtained almost absolute agreement between the paired analyses, reinforcing the repeatability of the algorithm.

The agreement at pixel level was also high. It obtained a SS of 0.78 and a Kappa coefficient of 0.76, which is considered a *substantial* agreement.

5.2.4. Study III – REP II

Objective:

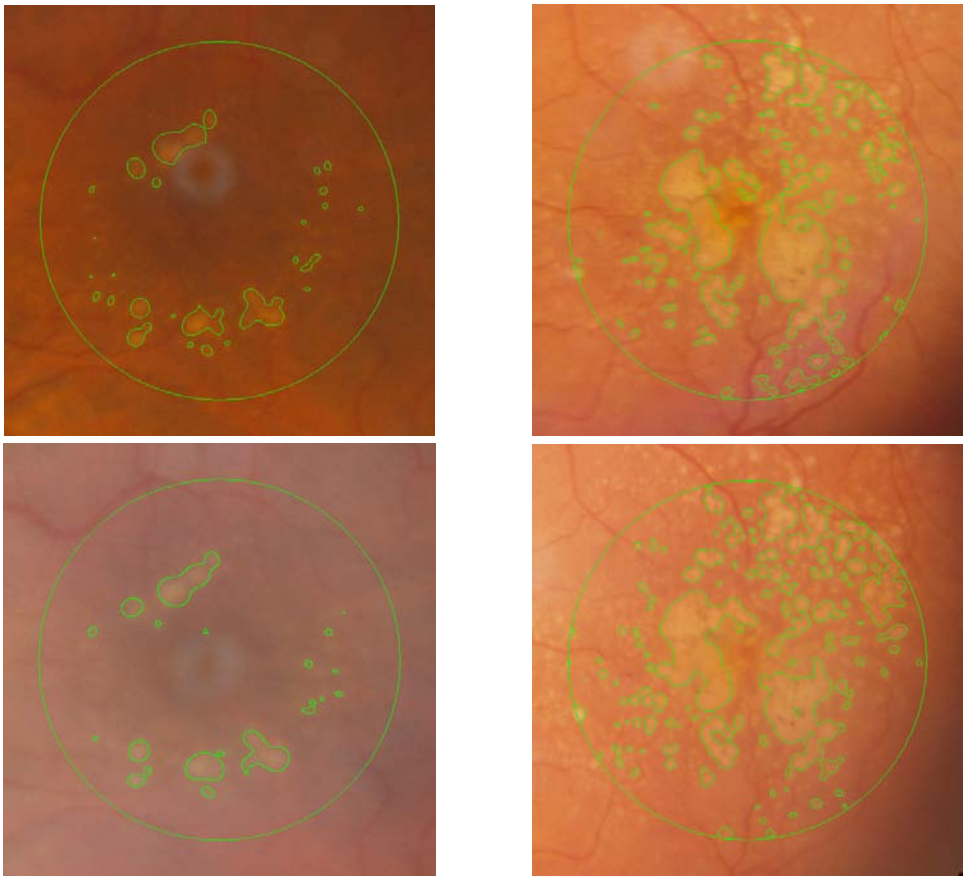
To evaluate the repeatability of the automated detection method to analyze images of one eye, before and after a cataract surgery.

Methods:

The reason for this study was the different origin and purpose of the images. In the previous study the images were acquired without clinical interventions in between. Here the images are from the ECAM study, one before and the second after the surgical intervention, in which it is expected to maintain the drusen morphology. Since the image quality changes between repetitions it is expected to have a higher variability and that is why the studies were separated.

The ECAM image dataset contained a total of 77 images. This was not an even number as it would be expected, since for some patients the acquisition was repeated at different stages of the recovery. From these were selected only the images which contained at least two images from the same retina and that were acquired in good conditions, which resulted on a total of 7 pairs of images (14 images). This last condition was determinant to exclude several images, because the images acquired before the surgical intervention were in many cases with high opacity and consequently smoothed and with low sharpness. Also there were many images that contained light flares which would be incorrectly identified as drusen by the software.

The procedure in this study consisted on analysing automatically each pair of images and evaluating the agreement between the two images (figure 5.26). After the drusen detection, the agreement was assessed using the areas comparison, *intra-class correlation coefficient*, *sensitivity* and *specificity* analysis and Cohen's *kappa* statistics, which all will be presented next.



#3 *#5*

Figure 5.26 – Examples of repeated images and their correspondent analysis.

Results:

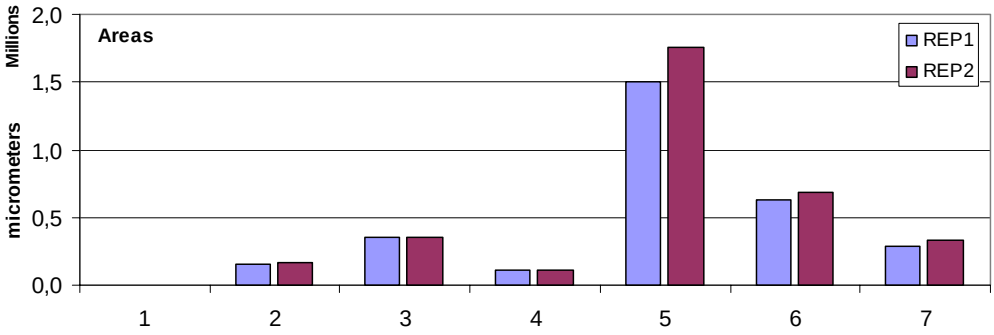
Outliers:

In this study it wasn't specified any outlier definition protocol, since the images which didn't have the required acquisition quality were previously removed from the dataset. In this sense none of the images should be classified as an outlier.

Areas:

The total area and the subfields areas defined in the Wisconsin grading system were calculated and displayed in the following figures and tables. In image #1 no drusen were detected by AD3RI therefore figure 5.27 has a value of 0 (zero).

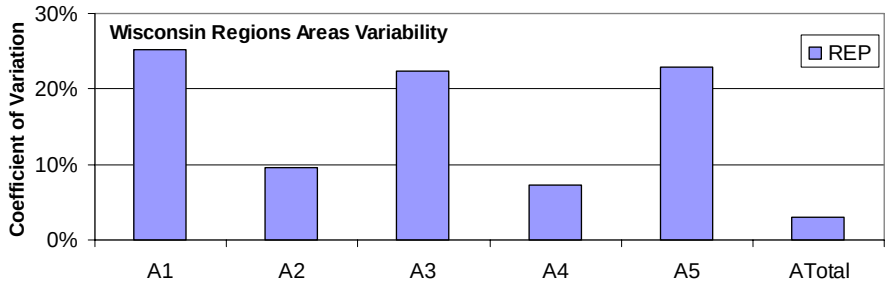
From figure 5.27 it can be observed that the areas are slightly higher for REP2 (after intervention). Nevertheless these are small increases that are not significant. The REP2 images have a better sharpness and for this reason have better defined contours and the smaller drusen are clearer, as well.



#	REP1	REP2	Mean	CV
1	0	0	0	0,0%
2	159835	162033	160934	0,7%
3	351161	354444	352803	0,5%
4	110653	112517	111585	0,8%
5	1506004	1757082	1631543	7,7%
6	626508	689598	658053	4,8%
7	286787	327338	307063	6,6%

Figure 5.27 – Graphical comparison of areas between the analysis 1 and 2 and total area details for each image.

When analysing the *coefficient of variability* along the five subfields of the Wisconsin circle and the whole area (figure 5.28), it is observed once again a high discrepancy between the Wisconsin subfields, but for the whole area the variability repeated the 3% of the previous study. As before, the variability among the Wisconsin subfields is due to differences on subfields with small affected areas.



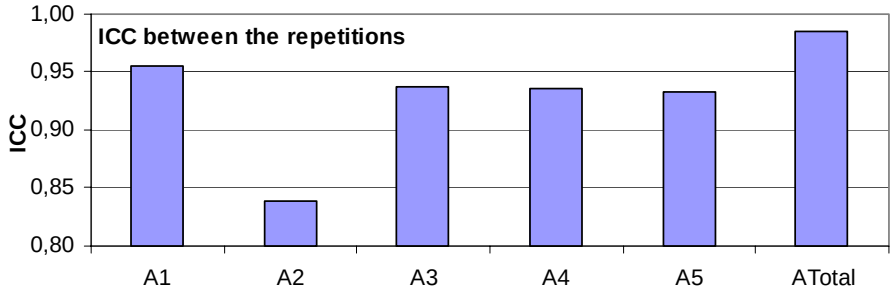
	A1	A2	A3	A4	A5	A1..A5	ATotal
REP	25%	10%	22%	7%	23%	17%	3%

Figure 5.28 – Mean area variability along the Wisconsin subfields and whole area.

Intra-Class Correlation Coefficient:

In this study, there was less consistency in the ICC in subfield A2. One of the main causes was image #5 (figure 5.26), which contains large affected areas, with small and large drusen, and where the AD3RI generated a visible difference on this subfield. However, this variability is influenced by the small size of this dataset, since one image represents 14% of the dataset. Therefore any difference provoked by one single image has direct influence on the

final result. Analysing the remaining subfields it can be seen that the agreement was high, the ICC was higher than 0.93 in all other Wisconsin subfields, and 0.98 on the total area.



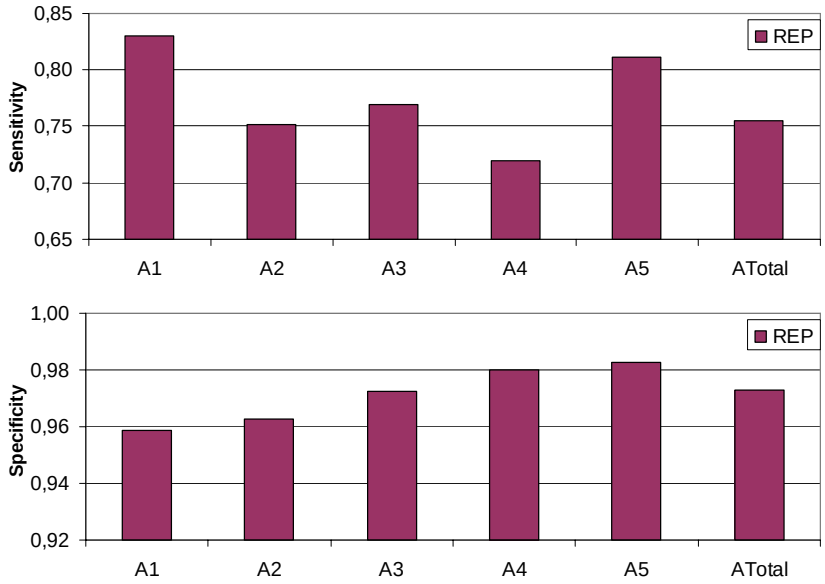
	A1	A2	A3	A4	A5	A1..A5	ATotal
REP	0,955	0,839	0,938	0,936	0,932	0,920	0,984

Figure 5.29 – Means ICC values in the Wisconsin subfields and the total area.

Sensitivity and specificity Analysis:

The *true positive*, *false positive*, *true negative* and *false negative* pixels were computed using the same procedure as in the previous study. The *sensitivity* (*SS*) and specificity (*SP*) results are presented in figure 5.30.

In the beginning of this study, it was expected to obtain less accurate results compared to the previous study (REP I). This was visible in the *SS* of the total area by decreasing from 0.78 to 0.75, but in the Wisconsin subfields it increased from 0.73 to 0.77 and the SP was almost unchanged. It is also visible that there are no significant differences among the SS and SP in the Wisconsin subfields.



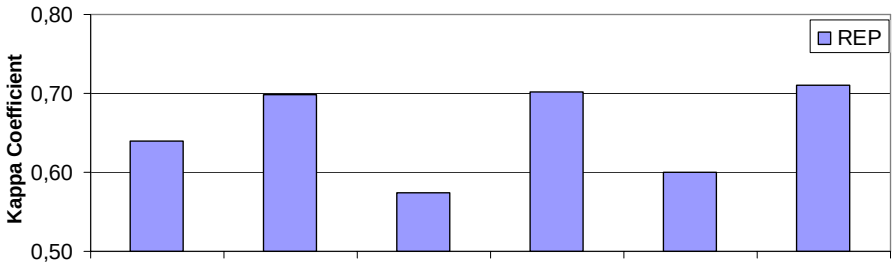
	Sensitivity							Specificity						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**	**A1**	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
REP	0,830	0,751	0,769	0,719	0,811	0,776	0,755	0,959	0,963	0,972	0,980	0,983	0,971	0,973

Figure 5.30 – *Sensitivity* and *Specificity* analysis in the Wisconsin subfields and total area.

Cohen's Kappa Statistics:

The Cohen's *kappa* coefficient was calculated using the same procedure as in the previous studies and obtained the results shown in figure 5.31. It should be noticed that the *kappa* coefficient obtains the same result despite the order of the images (REP1-REP2 or REP1-REP2) in opposition to the *SS* and the *SP*.

The general agreement was from *good* to *substantial* and it should be highlighted that when the image was analysed as a whole it obtained a consistent *substantial* agreement in both studies, although, the kappa coefficient decreased from the previous study from 0.76 to 0.71.



	Kappa Coefficient						
	A1	**A2**	**A3**	**A4**	**A5**	**A1..A5**	**ATotal**
REP	0,640	0,698	0,574	0,702	0,600	0,643	0,710

Figure 5.31 – *Kappa* Coefficient in the Wisconsin subfields and total area.

Conclusions:

As expected, the results from this repeatability study were slightly worst than the previous one, even though they were good results that can classify the AD3RI as a reliable tool. The mean error maintained the 3% average in the total area and increased to 17% when analysing the Wisconsin subfields. The ICC was also high (0.98) showing a high agreement between the paired images. As for the pixel level the agreement was good with a *kappa* coefficient of 0.71 and SS 0.755.

One of the major external factors that influenced negatively the results was the defects in the image acquisition. Many of the images from this dataset contain light artefacts, which influence the detection of drusen, namely light flares and lens opacity. In those images after the analyses the regions affected by light artefacts were manually cleaned from the *marked* images REP1 and REP2.

After cleaning the image acquisition artefacts, it can be said that the overall agreement in the quantification of drusen from repeated images is *substantial* and that the analysis criterion between the Wisconsin subfields is uniform.

5.3 - Summary

The studies now presented have successfully validated the automatic methodology for the detection of drusen. These included two important studies in which the methodology was compared to the markings performed manually by ophthalmologists and by trained graders. These were as important as difficult to obtain, since several other ophthalmologists were also invited to participate in the trials but unfortunately due to their professional duties they weren't been able to contribute with image markings. The two repeatability studies were also important to validate the capacity of AD3RI to evaluate drusen in sequences of images. It was necessary to divide into two studies, in order to analyse separately the two datasets, one more dedicated to repeatability and another to assess the reproducibility in images taken some months later.

The quality of the retinal images is one of the main problems for the automatic detection. A correct maintenance of the Retinographs is fundamental to keep the lenses as clear as possible and consequently avoid light flares. This fact unfortunately appeared in many images that were supplied. In these cases the algorithm was unable to differentiate those artefacts from drusen and needed a manual post-processing to remove those areas from all the specialists and AD3RI images.

The comparison with the specialists showed that AD3RI is able to accurately reproduce their drusen markings. In both studies (TRIAL I and TRIAL II) the AD3RI obtained equivalent or, in many cases better, classifications than the specialists obtained when compared among them, which indicates its ability to classify drusen.

The repeatability studies showed that the AD3RI is capable to repeat the analysis on images taken sequentially or taken some months later with an error margin of 3% on average. Although, more tests were required since these datasets contained only 16 images. The main problem that can compromise the results is the image acquisition quality. In some cases a manual user intervention still can fix it without decreasing the AD3RI accuracy, but in the majority of the cases the repeatability is compromised.

Since there is no standard assessment technique to be applied in these types of studies most of the works that are published use performance indicators that are not used by the others, making it difficult to compare the algorithms to each other. In this work the performance was assessed using various indicators, so that it can be compared with other algorithms.

The reproducibility measured using the *Coefficient of variation* was used by Sebag, Peli *et al.* (1991) in 17 images and obtained a mean value of 4.7%. Phillips, Forrester *et al.* (1993) assessed 13 images and obtained 17% to 5% range. In this study with a dataset of 16 images and without removing outliers it was obtained a mean value of 3%, which can be considered a good value taking in consideration that the images were not specifically prepared for this study, and had reasonable changes in the capturing conditions.

When assessing the *sensitivity*, *specificity* and *kappa* statistics, the results are equivalent or even better than the ones published. In the works of Smith, Chan *et al.* (2005) from a set of 20 images it was obtained a *sensitivity* of 0.70 and a *specificity* of 0.81, when comparing it to one expert grading. The AD3RI obtained in TRIAL I from 22 images and 8 experts a *sensitivity* of 0.69 and a *specificity* of 0.95. In TRIAL II where 2 experts graded 39 images, the AD3RI obtained even better results; 0.74 for *specificity* and 0.97 for *specificity*. The work of Soliz, Wilson *et al.* (2002) validated a set of 21 images which were analysed with an user-assisted procedure and compared to a ground truth analysis produced by trained graders from the Wisconsin Reading Center using guidelines found in the Wisconsin ARMD Grading Schema. The accuracy was assessed comparing the size of the bigger drusen found in each Wisconsin subfield. The agreement between the manual and the automated method obtained 0.712 by Cohen's *kappa* statistics. The AD3RI was assessed using a pixel-to-pixel agreement having obtained 0.577 and 0.605 in TRIAL I and TRIAL II, respectively. These values are lower than the Soliz's work, however their assessment technique is easier to achieve higher *kappa* values, for example if the area value is equal but identified in different locations, Soliz's will obtain an absolute agreement, while AD3RI will have null agreement, since there is no validation if the areas calculated are the same.

The weighed matching analysis was a novel method proposed to create a ground truth analyses when there were more than 2 specialists analysing the same images. It was only applicable to TRIAL I where it showed that the relative performance was maintained by a majority of the specialists, although for some, this change was visible. Another advantage is the performance due to its ability to be easily updated by including new gradings without the need to re-compare all the examples.

Chapter 6 -
Conclusions

This thesis presented a methodology for automatic detection and quantification of drusen deposits in retinal images using digital image processing techniques. The automation of such task allowed reproducible and accurate measurements to be made without the user intervention. This will improve the assessment of drusen progression, since the analysis will be performed using always the same criteria.

For Ophthalmologists the existence of such tool will contribute to a better understanding of drusen progression and early assessment of drusen changes, since the quantitative comparison of images taken at different occasions will be done more precisely and easily. This way they can evaluate if drusen are stable, progressing to other subfields or retracting.

The possibility to apply the algorithm in an unsupervised mode over large image datasets potentiate its use in epidemiologic studies and clinical trials which study retinal drusen, as it increases accuracy, repeatability and requires less human resources. It should be reminded that these latter are actually being made manually by trained grading specialists, which is fastidious and lacks repeatability. Many of these studies need to analyze thousands of images, which are graded by different specialists from different centres, which introduce a high variability in the results' analysis.

A literature review on automated drusen detection showed that most of the existing methods use threshold techniques to detect drusen, which have a tendency to produce false positives. Therefore, the strategy in this work was to exploit other image processing techniques to solve this problem. The methodology which was proposed was divided into four

steps: image pre-processing, drusen detection, drusen modelling and drusen quantification, and has been validated by comparison to the quantifications performed by ophthalmologists and trained graders which collaborated in this work.

For the image pre-processing step, it was needed to compensate the illumination which is usually deficient in retinal images acquisition. It was proposed an illumination correction algorithm based on smoothing splines fitting and contrast normalization. The combination of these two operations was an important step to obtain a less parameterized methodology, since they were able to deliver images with uniform illumination and normalized contrast to the subsequent steps.

The drusen detection algorithm which was presented, *Gradient Path Labelling*, is a novel segmentation method based on intensity gradient information that allows the detection and isolation of drusen spots even in low contrast images. Its comparison to the Watershed Transform, which is its most similar method, showed that it is able to exceed its performance in terms of complexity and execution time. The proposed GPL algorithm is able to maintain a linear complexity to the number of pixels, which can be important for large scale or high resolution problems.

To characterize drusen, since they exhibit a typical shape similar to a Gaussian function, a surface modelling approach with Gaussian functions was applied. This process begins by detecting spots to obtain initial function parameters and then a fitting algorithm to estimate the model. This technique exceeds the capabilities of the common segmentation algorithms by allowing a shape consistent segmentation. When the shape of the structures is known *a priori* this technique restricts the segmentation to the template shape. Also, with the model it is possible to extract drusen indications such as size, location, shape and eventually their volume estimation, which can be useful to evaluate disease progression.

To validate the proposed methodology two software prototypes were developed. For the manual drawing of drusen contours, the specialists mentioned above used the MD3RI application, which was developed specifically to assist the marking of images in a semi-automatic procedure. The automatic quantification process was fully automated in the AD3RI application, implemented in C++ language, the same as the MD3RI. The reason for choosing the C++ implementation was its better performance in comparison to the other programming languages, such as Java or Matlab.

The performance, in terms of the execution time, was a concern during this thesis. The objective was to create algorithms that could be easily introduced in medical assisting tools; therefore they were initially developed and afterwards optimized for performance.

To assess the accuracy of the methodology in quantifying drusen, four different studies were performed. Two of them were to assess the agreement between the specialists' analyses and the automated method, and another two to assess the repeatability of the automated method on two different datasets.

From this statistical analysis of the agreement with the specialists, it is concluded that the algorithm follows the same parameters as ophthalmologists, although more accurately. When compared to the specialists a first set of 22 images was analyzed by 8 specialists and the agreement was high, obtaining a *sensitivity* of 0.787 and a *kappa* coefficient of 0.7. These values were obtained using the proposed Weighed Matching Analysis algorithm which evaluates the similarity between one marked image and a set of images marked by the specialists. It is based on a pixel-to-pixel comparison of an image and a probability map which contains the statistical significant analyses of the specialists. The second set of 39 images involved two specialists and obtained a *sensitivity* of 0.741 and a *kappa* coefficient of 0.601, which can also be considered good results and confirm the previous ones.

The repeatability study was restricted to two datasets, containing 9 and 7 pairs of images, respectively. The agreement on the total area was high, achieving a *coefficient of variation* of 3% on both datasets. As for the pixel level the *sensitivity* obtained 0.78 and 0.75, and the *kappa* coefficient 0.76 and 0.71, respectively for dataset one and two. These values reveal a *substantial* agreement between repetitions. It should be noticed that these values were obtained with images which were not captured with the objective of assessing the software repeatability, being captured with different lightning and spaced in time.

As a general and critical analysis of this work, it can be said that the proposed methodology is able to detect and quantify drusen using an automated procedure, although the quality of the images have significant influence in its overall performance. The image pre-processing step, that is perhaps the most important step for quantification success, might not be able to correct all the lightning defects without affecting the drusen spots, especially in the presence of large drusen. Also, the manual confirmation of the normalization procedure could be improved in the future. Besides these limitations the automated drusen quantification methodology showed to be reliable and reproducible.

It should also be mentioned that the multi-disciplinary collaboration with the Ophthalmologists was extremely important to understand their difficulties to introduce these techniques in their daily activities and to analyse their technique of analysis.

The collaboration with the Hospital de Santa Maria was extended beyond this work, having developed a tool to measure retinal vessels' diameters and the installation of a digital image acquisition system in a non-mydriatic retinograph (Monteiro-Grillo, Lopes et al. 2006). This study objective was to evaluate veins to arteries ratio, in order to detect early signs of hypertension and other cardiac problems.

It should also be mentioned that this work was disseminated in several oral presentations, including the TV program *Magazine de Ciência e Tecnologia 2010* (Mora 2007), has published twelve articles in National and International Conferences (full list of publications on Annex C), has received from Fundação para Ciência e Tecnologia one grant to support the PhD candidate activities, and another to support a Research Project on this topic.

Chapter 7 -
Future Work

The problem of automatically extracting quantitative information from retinal images is still far from being a closed scientific case. The natural complexity of retinal images, the acquisition conditions many times far from being the ideal and the reduced image quality provoked by diseases of an aged eye, makes their automatic processing a true challenging task. Also, new diagnostic and imaging techniques are being developed to allow depth information, reflective and density characteristics and others that require specific tools for their analysis. The multi-modal data fusion between 2D imaging techniques, such as fundus or SLO images, and 3D techniques like OCT images, will certainly be a challenge to the development of new diagnostic techniques.

Another important step is to integrate these tools in the daily operations of an Ophthalmological Clinic. The human perception is still more accurate and comprehensive than the computer which simply analyses the data which the human asks it to do. This is due to the higher capacity of humans to correlate the data which they are observing than the computer. However, humans gets tired, specially to perform fastidious tasks which require constant precision. Computerized analysis tools should then be a compromise between automating the analyses and helping the clinicians to perform their own medical diagnosis. The tool which was presented is one of such applications; although, until now and from what is of the author's knowledge, there are still no proven tools to automate quantification and being used in clinical practice.

The work presented in this thesis only exploited the detection of drusen spots in images correctly acquired and evincing drusen. These requisites limited the number of cases

which can be treated by this methodology. The adaptation of the algorithms to identify drusen and other structures from images containing different pathologies would be the next development. The use of images colour information and the extraction of other features from the spots such as dimensions, shape factors, colour, location, intensity and others could be inputs of an automatic classifier which identifies the type of spots being analysed (dry and wet drusen, cotton wool spots, exudates, tumours and others).

The image pre-processing has still some issues to solve, especially in the case where the retina contains large and abnormal pathologies (tumours, cataracts and others). Also, the correction or elimination of light flares and other lightening artefacts would improve the image analysis. The identification in the image of light flares is not an easy task. They are characterized by a round or doughnut shape and by being predominantly white. The methodology needs to detect these artefacts and eliminate these areas to fix the illumination in those regions. Another solution would be to acquire several sequenced images with small eye movements, register the images and get the median image (each pixel is the median of the several observations).

Nevertheless a good quality image is always the best source for any automatic image processing tool. In order to improve the quality of the images used, the specification of a protocol for the image acquisition, that could guarantee only good quality images, is also important.

There are few published works on retinal image registration and it is an important step to compare images taken in different occasions and other imaging modalities. A method to automatically extract the vessels from the image and identify the referencing points that would be used in the register transformation, is still an open field. There are published works on vessels detection (Staal, Abràmoff et al. 2004; Oloumi, Rangayyan et al. 2007; Rezatofighi, Roodaki et al. 2008) but apparently none of them is focused on the registration problem, instead they verify the presence of aneurisms.

The improvement of the modelling procedure can also be a future work. On the one hand to reduce the computation time it could be interesting to evaluate its performance when deployed to run on the computers' GPU, since nowadays these are faster processors to compute intensive image operations. On the other hand a multi-stage analysis could improve the results by refitting the model with more functions, in particular, in the areas with medium or larger errors.

Annex A -

Generalized Gaussian Function

To model the drusen spots, Generalized Gaussian functions were used. The Levenberg-Marquardt optimization algorithm was applied to find the optimal parameters which adjust the Gaussian to the drusen image. This algorithm requires the function's Jacobian matrix, i.e., the partial derivatives for every adjustable parameter in every vector points, to be calculated. Both analytical and finite difference approximated Jacobian could be used in the algorithm. However, since the function is known *a priori*, an analytical calculation was used.

The Generalized Gaussian function used was:

$$G(x,y) = A \cdot e^{-0.5*\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x \cdot s_f}\right)^2\right)^{\frac{2}{\beta}}} + z_0$$

where

$$X = (x - x_0)\cdot\cos(\theta) + (y - y_0)\cdot\sin(\theta)$$
$$Y = -(x - x_0)\cdot\sin(\theta) + (y - y_0)\cdot\cos(\theta)$$

The partial analytical derivatives with respect to the adjustable parameters were calculated using Mathcad (2007), being presented below.

Shifting parameters:

$$\frac{dG(x,y)}{dx_0} = \frac{G(x,y)\cdot\left(\frac{2*\cos\theta*X}{{\sigma_x}^2}+\frac{2*\sin\theta*Y}{{\sigma_x}^2\cdot{s_f}^2}\right)\cdot\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}-1}}{\beta}$$

$$\frac{dG(x,y)}{dy_0} = \frac{G(x,y)\cdot\left(-\frac{2*\sin\theta*X}{\sigma_x{}^2}+\frac{2*\cos\theta*Y}{\sigma_x{}^2\cdot s_f{}^2}\right)\cdot\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}-1}}{\beta}$$

$$\frac{dG(x,y)}{dZ_0} = 1$$

Amplitude parameter:

$$\frac{dG(x,y)}{dA} = e^{-0.5*\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}}}$$

Width parameters:

$$\frac{dG(x,y)}{d\sigma_x} = \frac{G(x,y)\cdot\left(\frac{2}{\sigma_x}\cdot\left(\frac{X}{\sigma_x}\right)^2+\frac{2}{\sigma_x}\cdot\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)\cdot\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}-1}}{\beta}$$

$$\frac{dG(x,y)}{ds_f} = \frac{G(x,y)\cdot\left(2\cdot Y^2\right)\cdot\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}-1}}{\beta\cdot\sigma_x{}^2\cdot s_f{}^3}$$

Profile parameter:

$$\frac{dG(x,y)}{d\beta} = \frac{G(x,y)\cdot\ln\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)\cdot\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}}}{\beta^2}$$

Rotation parameters:

$$\frac{dG(x,y)}{d\theta} = \frac{G(x,y)\cdot\left(\left(\frac{2\cdot X\cdot Y}{\sigma_x{}^2}\right)-\left(\frac{2\cdot X\cdot Y}{\sigma_x{}^2\cdot s_f{}^2}\right)\right)\cdot\left(\left(\frac{X}{\sigma_x}\right)^2+\left(\frac{Y}{\sigma_x\cdot s_f}\right)^2\right)^{\frac{2}{\beta}-1}}{\beta^2}$$

Annex B –

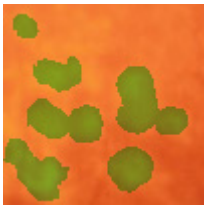
Image Grading Protocol

PROTOCOL

Manual Drusen Deposits Detection on Retina Images

Rev. 1.1

February 2010



MD3RI
Manual Drusen Deposits Detection on Retinal Images

PROTOCOL

Document date: 17 Feb 2010

Drusen (definition):

- Drusen are deposits of extracellular material beneath the retina surface causing it to rise.
- They appear in fundus images as yellow round shaped spots with a diameter higher 63µm. Use Tool *Min Size (F7)* to show minimum druse size over the image.
- Their contours are smoothed as consequence of being located beneath a membrane, which will be raised with a smooth slope. This characteristic is main difficult for achieving a precise and reproducible quantification of drusen areas.
- The minimum intensity for a spot to be considered as druse can vary. The criteria should be defined by the user in terms of contrast with the background. Typical contrast would be at least 15 to 20 intensity levels between druse and background. (see intensity below (x,y) coordinates)

Example:

Original *Marked*

Procedure:

- Apply a mean filter to the image before running the detection algorithm
- Use the Green image
- Mark the drusens using the semi-automatic and retouch them using the manual mode when needed
- Save the analysis (*Save Analysis* button)
- (optional) Save the marked image (*Save Marked* button)
- Write any comment you might think is important to justify your analysis in the field *Observations*.

Notes:

- Keep the initial ROI location
- Try to use illumination and contrast to a minimum

Annex C – List of Publications

Mora, A., P. Vieira, et al. (2004). Drusen Deposits on Retina Images: Detection and Modeling. in proceedings of MEDSIP-2004, Malta.

Mora, A., J. Fonseca, et al. (2005). Drusen Deposits Modelling with Illumination Correction. in proceedings of Biomed-2005, Innsbruck, Austria.

Mora, A., P. Vieira, et al. (2005). Modeling of Drusen Deposits Based on Retina Image Tridimensional Information. in proceedings of Second International Conference on Computacional Intelligence in Medicine and Healthcare - CIMED-2005, Costa da Caparica, Portugal.

Fonseca, J., A. Mora, et al. (2005). Detecção de Drusas em imagens de retinografia. in proceedings of 3ª Jornadas de Engenharia de Electrónica e Telecomunicações e de Computadores, ISEL, Lisboa, Portugal.

Mora, A., P. Vieira, et al. (2006). MD3RI a Tool for Computer-Aided Drusens Contour Drawing. in proceedings of Fourth IASTED International Conference on Biomedical Engineering - BIOMED2006, Innsbruck, Austria, ACTA Press.

Mora, A., J. Fonseca, et al. (2006). Quantification of Drusen Deposits using Image Processing Techniques. in proceedings of European Association for Vison and Eye Research - EVER2006.

Moitinho, F., A. Mora, et al. (2006). AD3RI a Tool for Computer - Automatic Drusen Detection. in proceedings of CompIMAGE - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications, Coimbra, Portugal.

Moitinho, F., A. Mora, et al. (2007). A Drusen volume quantification method based on a segmentation algorithm. in proceedings of VIP Image - Thematic Conference on Computational Vision and Medical Image Processing, Porto, Portugal.

Moitinho, F., A. Mora, et al. (2007). Image Segmentation for Drusen Spots Detection and Modelling. in proceedings of CIMED2007 - Third International Conference on Computational Intelligence in Medicine and Healthcare, Plymouth, England.

Mora, A., J. Fonseca, et al. (2009). Retina Image Gradings' Comparison by Weighted Matching Analysis. in proceedings of World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, Springer.

Mora, A., P. Vieira, et al. (2010). "Advances in Image Processing Techniques for Drusens Detection and Quantification in Fundus Images." Emerging Trends in Technological Innovation: 299-307.

Mora, A. D., P. M. Vieira, et al. (2010). The Weighed Matching Analysis method and the accuracy of automated retinal image processing. in proceedings of IBERSENSOR - 7TH Ibero-American Congress on Sensors, Lisboa.

Bibliography

Aberdeen-Biomedical-Imaging-Centre. (2009). "SLO - Scanning Laser Opthalmoscope",Retrieved 19-07-2009, 2009, from http://www.abdn.ac.uk/ims/imaging/facilities/slo/index.php.

AMD Alliance International (2006). What's Happening in AMD? An answer based on presentations at AAO's "Retina 2006: Emerging New Concepts" - Internal Report, AMD Alliance International.

American Health Assistance Foundation. (2009). "Medical Illustrations",Retrieved 29-07-2009, from http://www.ahaf.org/macular/about/understanding/medical_illustrations.html.

Amsler Grid (2010). Amsler Grid.

AREDS - National Eye Institute. (2009). "The Age-Related Eye Disease Study 2 ",Retrieved 20-07-2009, from http://www.areds2.org/.

Atchison, D., N. Pritchard, et al. (2005). "Shape of the retinal surface in emmetropia and myopia." Investigative ophthalmology & visual science 46(8): 2698.

Attene, M., B. Falcidieno, et al. (2006). "Hierarchical mesh segmentation based on fitting primitives." The Visual Computer 22(3): 181-193.

Augood, C. A., J. R. Vingerling, et al. (2006). "Prevalence of age-related maculopathy in older Europeans: the European Eye Study (EUREYE)." Arch Ophthalmol 124(4): 529-535.

Baeg, M., H. Hashimoto, et al. (1995). Pose estimation of quadratic surface using surface fitting technique. in proceedings of International Conference on Intelligent Robots and Systems.

Beucher, S. and C. Lantuejoul (1979). Use of watersheds in contour detection. in proceedings of International Workshop on image processing, real-time edge and motion detection/estimation, Rennes, France.

Biegelbauer, G. and M. Vincze (2007). Efficient 3D object detection by fitting superquadrics to range image data for robot's object manipulation. in proceedings of IEEE International Conference on Robotics and Automation.

Bird, A. C., N. M. Bressler, et al. (1995). "An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group." Surv Ophthalmol 39(5): 367-374.

Björck, A. and Å. Björck (1996). Numerical methods for least squares problems, Society for Industrial Mathematics.

Bland, M. and D. Altman (1986). "Statistical methods for assessing agreement between two methods of clinical measurement." The Lancet 327(8476): 307-310.

Blanz, V., K. Scherbaum, et al. (2007). Fitting a morphable model to 3D scans of faces. in proceedings of International Conference on Computer Vision - ICCV.

Borland Software Corporation (2002). Borland C++Builder v6.

Brandon, L. (2003). Automated Drusen Detection in a Retinal Image Using Multi-level Analysis. Master of Science Thesis, Clemson University

Brandon, L. and A. Hoover (2003). Drusen Detection in a Retinal Image Using Multi-level Analysis. in proceedings of Medical Image Computing and Computer-Assisted Intervention.

Bressler, S. B., B. Munoz, et al. (2008). "Racial differences in the prevalence of age-related macular degeneration: the Salisbury Eye Evaluation (SEE) Project." Arch Ophthalmol 126(2): 241-245.

Brunner, S., I. Krebs, et al. (2005). Cataract Surgery in Nonexsudative Age-Related Macular - First Results of a Prospective, Randomized, Multicenter Trial (ECAM-1). in proceedings of Association for Research in Vision and Ophthalmology.

Buch, H., N. V. Nielsen, et al. (2005). "14-year incidence, progression, and visual morbidity of age-related maculopathy: the Copenhagen City Eye Study." Ophthalmology 112(5): 787-798.

Canon. (2009). "Canon – Healthcare Products",Retrieved 19-07-2009, from http://www.usa.canon.com/html/industrial_medeq/ophthalmic_index.html.

Chan, T. and L. Vese (2001). "Active contours without edges." IEEE Transactions on image processing 10(2): 266-277.

Chen, S. J., C. Y. Cheng, et al. (2008). "Prevalence and associated risk factors of age-related macular degeneration in an elderly Chinese population in Taiwan: the Shihpai Eye Study." Invest Ophthalmol Vis Sci 49(7): 3126-3133.

Cohen, J. (1960). "A coefficient of agreement for nominal scales." EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT 20(1): 37.

Cook, H. L., P. J. Patel, et al. (2008). "Age-related macular degeneration: diagnosis and management." Br Med Bull 85: 127-149.

Cormen, T. H., C. E. Leiserson, et al. (2001). Introduction to algorithms, The MIT press.

Cukras, C. and F. Ferris (2008). "Cardiovascular Risk Factors and Age-Related Macular Degeneration: the Los Angeles Latino Eye Study." Evidence-Based Ophthalmology 9(3): 193.

Davidovits, P. (2001). Physics in Biology and Medicine - Second Edition, Harcourt / Academic Press, 215.

De Jong, K. (1975). An analysis of the behavior of a class of genetic adaptive systems, University Microfilms International.

De Sousa Meneses, D., G. Gruener, et al. (2005). "Causal Voigt profile for modeling reflectivity spectra of glasses." Journal of Non-Crystalline Solids 351(2): 124-129.

Delcourt, C., J. L. Diaz, et al. (1998). "Smoking and age-related macular degeneration. The POLA Study. Pathologies Oculaires Liees a l'Age." Arch Ophthalmol 116(8): 1031-1035.

Delori, F. and K. Pflibsen (1989). "Spectral reflectance of the human ocular fundus." Applied Optics 28(6): 1061-1077.

Donoso, L., D. Kim, et al. (2006). "The role of inflammation in the pathogenesis of age-related macular degeneration." Survey of ophthalmology 51(2): 137-152.

Driggers, R. (2003). Encyclopedia of optical engineering, CRC Press.

Duanggate, C. and B. Uyyanonvara (2008). A Review of Automatic Drusen Detection and Segmentation from Retinal Images. *in proceedings of* 3rd International Symposium on Biomedical Engineering.

Duda, R. and P. Hart (1973). Pattern Classification and Scene Analysis, John Wiley and Sons**:** 271-272.

Enzmann, V., E. Yolcu, et al. (2009). "Stem cells as tools in regenerative therapy for retinal degeneration." Archives of ophthalmology **127**(4): 563.

Fercher, A. F. (1990). Ophthalmic interferometry. *in proceedings of* International Conference on Optics in Life Sciences, Garmisch-Partenkirchen, Germany.

Ferris, F. L., M. D. Davis, et al. (2005). "A simplified severity scale for age-related macular degeneration: AREDS Report No. 18." Arch Ophthalmol **123**(11): 1570-1574.

Fine, A., M. Elman, et al. (1986). "Earliest symptoms caused by neovascular membranes in the macula." Archives of ophthalmology **104**(4): 513.

Finlayson, G. D., B. Schiele, et al. (1998). Comprehensive Colour Image Normalization. Computer Vision — ECCV'98**:** 475.

Fisher, S. (1970). Statistical methods for research workers, Oliver and Boyd.

Fletcher, A., G. Bentham, et al. (2008). "Sunlight exposure, antioxidants, and age-related macular degeneration." Archives of ophthalmology **126**(10): 1396.

Frank, R., J. Puklin, et al. (2000). "Race, iris color, and age-related macular degeneration." Transactions of the American Ophthalmological Society **98**: 109.

Friedman, D. S., J. Katz, et al. (1999). "Racial differences in the prevalence of age-related macular degeneration: the Baltimore Eye Survey." Ophthalmology **106**(6): 1049-1055.

Friedman, D. S., B. J. O'Colmain, et al. (2004). "Prevalence of age-related macular degeneration in the United States." Arch Ophthalmol **122**(4): 564-572.

Goatman, K. A., A. D. Whitwam, et al. (2003). Colour normalisation of retinal images. *in proceedings of* Medical Image Understanding and Analysis, Sheffield.

Gonzalez, R. and R. Woods (2007). Digital Image Processing - third edition, Prentice-Wall.

Gulsen, M., A. Smith, et al. (1995). "A genetic algorithm approach to curve fitting." International Journal of Production Research **33**(7): 1911-1924.

Haddad, S., C. Chen, et al. (2006). "The genetics of age-related macular degeneration: a review of progress to date." Survey of ophthalmology **51**(4): 316-363.

Hageman, G. S., P. J. Luthert, et al. (2001). "An integrated hypothesis that considers drusen as biomarkers of immune-mediated processes at the RPE-Bruch's membrane interface in aging and age-related macular degeneration." Prog Retin Eye Res **20**(6): 705-732.

Hagyard, D., M. Razaz, et al. (1996). Analysis of watershed algorithms for greyscale images. *in proceedings of* Int. Conf. on Image Processing.

Helmholtz, H. v. (1851). The Description of an Ophthalmoscope, Cleveland press.

High Display. (2010). "Monitor Calibration",Retrieved 10-05-2010, *from* http://www.highdisplay.com/monitor-calibration/.

Hogg, R. and P. Andersen (2006) "Quantum-dot diodes provide sources for optical coherence tomography." SPIE Newsroom.

Holland, J. (1975). "Adaptation in natural and artificial systems." Ann Arbor MI: University of Michigan Press.

Hoover, A. and M. Goldbaum (2001). Illumination equalization of a retinal image using the blood vessels as a reference. *in proceedings of* Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO).

Hoover, A. and M. Goldbaum (2003). "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels." IEEE Trans Med Imaging **22**(8): 951-958.

Hubbard, L. D., R. P. Danis, et al. (2008). "Brightness, contrast, and color balance of digital versus film retinal images in the age-related eye disease study 2." Invest Ophthalmol Vis Sci **49**(8): 3269-3282.

Jackowski, M., M. Satter, et al. (2003). "Approximating digital 3D shapes by rational Gaussian surfaces." Visualization and Computer Graphics, IEEE Transactions on **9**(1): 56-69.

Jagoe, J. R., C. I. Blauth, et al. (1990). "Quantification of retinal damage during cardiopulmonary bypass: comparison of computer and human assessment." Communications, Speech and Vision, IEE Proceedings I **137**(3): 170-175.

Johnson, P. T., G. P. Lewis, et al. (2003). "Drusen-associated degeneration in the retina." Invest Ophthalmol Vis Sci **44**(10): 4481-4488.

Jonas, J. B., G. C. Gusek, et al. (1988). "Variability of the real dimensions of normal human optic discs." Graefes Arch Clin Exp Ophthalmol **226**(4): 332-336.

Jonasson, F., A. Arnarsson, et al. (2003). "The prevalence of age-related maculopathy in iceland: Reykjavik eye study." Arch Ophthalmol **121**(3): 379-385.

Khanifar, A. A., A. F. Koreishi, et al. (2008). "Drusen ultrastructure imaging with spectral domain optical coherence tomography in age-related macular degeneration." Ophthalmology **115**(11): 1883-1890.

Kirkpatrick, J. N. P., T. Spencer, et al. (1995). "Quantitative image analysis of macular drusen from fundus photographs and scanning laser ophthalmoscope images." Eye (Royal College of Ophthalmologists) **9**: 48-55.

Klein, R., M. D. Davis, et al. (1991). "The Wisconsin age-related maculopathy grading system." Ophthalmology **98**(7): 1128-1134.

Klein, R., B. E. Klein, et al. (2002). "Ten-year incidence and progression of age-related maculopathy: The Beaver Dam eye study." Ophthalmology **109**(10): 1767-1779.

Kruiskamp, W. and D. Leenaerts (1995). DARWIN: CMOS opamp synthesis by means of a genetic algorithm. *in proceedings of* Conf. Design Automation, ACM New York, NY, USA.

Krupinski, E., R. Weinstein, et al. (1996). "Experience-related differences in diagnosis from medical images displayed on monitors." Telemedicine Journal **2**(2): 101-108.

Landis, J. and G. Koch (1977). "The measurement of observer agreement for categorical data." Biometrics: 159-174.

Leandro, J. J. G., R. M. Cesar, et al. (2001). "Blood vessels segmentation in retina: Preliminary assessment of the mathematical morphology and of the wavelet transform techniques." Xiv Brazilian Symposium on Computer Graphics and Image Processing, Proceedings: 84-90.

Lester, E. D. (1998). Feature extraction, image segmentation, and surface fitting: the development of a 3D scene reconstruction system. University of Tennessee

Levenberg, K. (1944). "A method for the solution of certain non-linear problems in least squares." Q. Appl. Math. **2**: 164-168.

Lim, C. and E. Sim (2005). Production planning in manufacturing/remanufacturing environment using genetic algorithm. *in proceedings of* Genetic And Evolutionary Computation Conference ACM.

Lim, J. I., L. LaBree, et al. (2000). "A comparison of digital nonmydriatic fundus imaging with standard 35-millimeter slides for diabetic retinopathy." Ophthalmology **107**(5): 866-870.

Liverpool Ocular Oncology Centre. (2005). "Eye Cancer - Royal Liverpool Ocular Oncology Centre",Retrieved 29-07-2009, *from* http://www.eyetumour.com/ocular_tumours.php.

Machemer, R., G. Michelson, et al. (2009). "The Atlas of Ophthalmology ",Retrieved 31-07-2009, *from* http://www.atlasophthalmology.com/.

Manivannan, A., J. N. Kirkpatrick, et al. (1998). "Novel approach towards colour imaging using a scanning laser ophthalmoscope." Br J Ophthalmol **82**(4): 342-345.

Manivannan, A., P. F. Sharp, et al. (1993). "Digital fundus imaging using a scanning laser ophthalmoscope." Physiol Meas **14**(1): 43-56.

Marquardt, D. W. (1963). "An algorithm for least-squares estimation of non-linear parameters." Journal of the Society for Industrial and Applied Mathematics **11**(2): 431-441.

Mathcad (2007). Mathcad - Engineering calculation software.

MathWorks (2002). MATLAB - The Language Of Technical Computing.

McCarty, C. A., B. N. Mukesh, et al. (2001). "Risk factors for age-related maculopathy: the Visual Impairment Project." Arch Ophthalmol **119**(10): 1455-1462.

McGraw, K. O. and S. P. Wong (1996). "Forming inferences about some intraclass correlation coefficients." Psychological Methods **1**(1): 30-46.

Meyer, F. and S. Beucher (1990). "Morphological segmentation." Journal of Visual Communication and Image Representation **1**(1): 21-46.

Miller, J. (2010). "Treatment of age-related macular degeneration: Beyond VEGF." Japanese Journal of Ophthalmology **54**(6): 523-528.

Miller, K. R. (2006). "Indocyanine Green Angiography",Retrieved 19-07-2009, *from* http://www.opsweb.org/OpPhoto/Angio/ICG/ICG.html.

Moitinho, F. M. B. (2009). Deteção Automática de Drusas em Imagens de Retinografia. Master of Science Thesis, FCT-UNL

Monteiro-Grillo, M., D. Lopes, et al. (2006). Software for the semi-automatic determination of the retinal vessel diameter. *in proceedings of* EVER-2006, Vilamoura, Portugal.

Mora, A. (2002). Automatização do Processo de Estimação do Estado de Conservação de Materiais Utilizando Espectrometria Laser. Master of Science Thesis,

Mora, A. (2007). 2010 - Magazine de Ciência e Tecnologia. Vasco Trigo, RTP2.

Mora, A., J. Fonseca, et al. (2009). Retina Image Gradings' Comparison by Weighted Matching Analysis. *in proceedings of* World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, Springer.

Mora, A., P. Vieira, et al. (2006). MD3RI a Tool for Computer-Aided Drusens Contour Drawing. *in proceedings of* Fourth IASTED International Conference on Biomedical Engineering - BIOMED2006, Innsbruck, Austria, ACTA Press.

Moreira, C. (2008). Sistema automático de identificação e caracterização de zonas de actividade em imagens solares. Master of Science Thesis, FCT-UNL

Morgan, W. H., R. L. Cooper, et al. (1994). "Automated extraction and quantification of macular drusen from fundal photographs." Australian and New Zealand Journal of Ophthalmology **22**(1): 7-12.

Narasimha-Iyer, H., A. Can, et al. (2006). "Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy." IEEE Transactions on Biomedical Engineering **53**(6): 1084-1098.

National Eye Institute. (2009). "Photos, Images, and Videos",Retrieved 28-07-2009, *from* http://www.nei.nih.gov/photo/.

Nichols, D. P. (1998) "CHOOSING AN INTRACLASS CORRELATION COEFFICIENT." SPSS Keywords.

Oloumi, F., R. M. Rangayyan, et al. (2007). "Detection of blood vessels in fundus images of the retina using Gabor wavelets." Conf Proc IEEE Eng Med Biol Soc **2007**: 6452-6455.

Ooi, C. and P. Tan (2003). "Genetic algorithms applied to multi-class prediction for the analysis of gene expression data." Bioinformatics **19**(1): 37.

Osareh, A., M. Mirmehdi, et al. (2002). Classification and Localisation of Diabetic-Related Eye Disease. Computer Vision — ECCV 2002**:** 325-329.

Oshima, Y., T. Ishibashi, et al. (2001). "Prevalence of age related maculopathy in a representative Japanese population: the Hisayama study." Br J Ophthalmol **85**(10): 1153-1157.

Otsu, N. (1975). "A threshold selection method from gray-level histograms." Automatica **11**: 285-296.

Owens, S. L. (1996). "Indocyanine green angiography." British Journal of Ophthalmology **80**(3): 263-266.

Pagliarini, S., A. Moramarco, et al. (1997). "Age-related macular disease in rural southern Italy." Arch Ophthalmol **115**(5): 616-622.

Pauleikhoff, D., M. J. Barondes, et al. (1990). "Drusen as risk factors in age-related macular disease." Am J Ophthalmol **109**(1): 38-43.

Peli, E. (1990). "Contrast in complex images." Journal of the Optical Society of America A **7**(10): 2032-2040.

Peli, E. and M. Lahav (1986). "Drusen Measurement from Fundus Photographs Using Computer Image Analysis." Ophtalmology **93**, December, 1986, 1575-1580.

Phillips, R., J. Forrester, et al. (1993). "Automated detection and quantification of retinal exudates." Graefes Arch Clin Exp Ophthalmol **231**(2): 90-94.

Phillips, R. P., T. Spencer, et al. (1991). "Quantification of diabetic maculopathy by digital imaging of the fundus." Eye **5 (Pt 1)**: 130-137.

Press, W., S. Teukolsky, et al. (1992). Numerical recipes in C, Cambridge university press Cambridge.

Prevent Blindness America (2008). Vision Problems in the U.S.: Prevalence of Adult Vision Impairment and Age-Related Eye Disease in America, update to the 4th ed. Prevent Blindness America.

Rapantzikos, K., M. Zervakis, et al. (2003). "Detection and segmentation of drusen deposits on human retina: potential in the diagnosis of age-related macular degeneration." Med Image Anal **7**(1): 95-108.

Rezatofighi, S. H., A. Roodaki, et al. (2008). "An enhanced segmentation of blood vessels in retinal images using contourlet." Conf Proc IEEE Eng Med Biol Soc **2008**: 3530-3533.

Rodgers, J. and W. Nicewander (1988). "Thirteen ways to look at the correlation coefficient." American Statistician **42**(1): 59-66.

Roerdink, J. B. T. M. and A. Meijster (2000). "The Watershed Transform: Definitions, Algorithms and Parallelization Techniques." Fundamenta Informaticae - IOS Press **41**: 187-228.

Rosenfeld, A. and A. C. Kak (1982). Digital picture processing, Academic Press, Inc. Orlando, FL, USA.

Sandberg, M., A. Weiner, et al. (1998). "High-risk characteristics of fellow eyes of patients with unilateral neovascular age-related macular degeneration." Ophthalmology **105**(3): 441-447.

Sarks, J., S. Sarks, et al. (1994). "Evolution of soft drusen in age-related macular degeneration." Eye (London, England) **8**: 269.

Savol, A. M., C. C. Li, et al. (1980). "Computer-aided recognition of small rounded pneumoconiosis opacities in chest x-rays." IEEE Transactions on Pattern Analysis and Machine Intelligence **2**(5): pp. 479-482.

Sbeh, Z. B., L. D. Cohen, et al. (2001). "A New Approach of Geodesic Reconstruction for Drusen Segmentation in Eye Fundus Images." IEEE TRANSACTIONS ON MEDICAL IMAGING **20**(12): 1321-1333.

Sbeh, Z. B., L. D. Cohen, et al. (1997). An adaptive contrast method for segmentation of drusen. *in proceedings of* International Conference on Image Processing.

Schachat, A. P., L. Hyman, et al. (1995). "Features of age-related macular degeneration in a black population. The Barbados Eye Study Group." Arch Ophthalmol **113**(6): 728-735.

Scheider, A. and L. Neuhauser (1992). "Fluorescence characteristics of drusen during indocyanine-green angiography and their possible correlation with choroidal perfusion." Ger J Ophthalmol **1**(5): 328-334.

Scholl, H., T. Peto, et al. (2003). "Inter-and intra-observer variability in grading lesions of age-related maculopathy and macular degeneration." Graefe's Archive for Clinical and Experimental Ophthalmology **241**(1): 39-47.

Schweitzer, D., C. Schrödel, et al. (1985). "Reflectance spectrophotometry in the human ocular fundus." Graefe's Archive for Clinical and Experimental Ophthalmology **223**(4): 207-210.

Sebag, M., E. Peli, et al. (1991). "Image analysis of changes in drusen area." Acta Ophtalmologica **69**: 603-610.

Seddon, J., S. George, et al. (2006). "Cigarette smoking, fish consumption, omega-3 fatty acid intake, and associations with age-related macular degeneration: the US Twin Study of Age-Related Macular Degeneration." Archives of ophthalmology **124**(7): 995.

Sedgewick, J. (2008). Scientific imaging wth Photoshop: methods, measurement, and output, Peachpit Press.

Sharp, P. F. and A. Manivannan (1997). "The scanning laser ophthalmoscope." Physics in Medicine and Biology **42**(5): 951-966.

Shin, D., N. Javornik, et al. (1999). "Computer-assisted, interactive fundus image processing for macular drusen quantitation." Ophthalmology **106**(6): 1119-1125.

Sinthanayothin, C. (1999). Image Analysis for Automatic Diagnosis of Diabetic Retinopathy. PhD Thesis, University of London

Sinthanayothin, C., V. Kongbunkiat, et al. (2003). Automated Screening System for Diabetic Retinopathy. *in proceedings of* ISPA03, Roma, Italy.

Skenduli-Bala, E., S. de Voogd, et al. (2005). "Causes of incident visual field loss in a general elderly population: the Rotterdam study." Arch Ophthalmol **123**(2): 233-238.

Smith, R. T., J. K. Chan, et al. (2005). "Automated detection of macular drusen using geometric background leveling and threshold selection." Arch Ophthalmol **123**(2): 200-206.

Smith, R. T., J. K. Chan, et al. (2005). "A method of drusen measurement based on reconstruction of fundus background reflectance." Br J Ophthalmol **89**(1): 87-91.

Smith, W., J. Assink, et al. (2001). "Risk factors for age-related macular degeneration:: Pooled findings from three continents." Ophthalmology **108**(4): 697-704.

Soliz, P., M. P. Wilson, et al. (2002). Computer-aided methods for quantitative assessment of longitudinal changes in retinal images presenting with maculopathy. *in proceedings of* Medical Imaging 2002: Visualization, Image-Guided Procedures, and Display, San Diego, CA, USA, SPIE.

Song, M., R. Haralick, et al. (2002). "Integrated surface model optimization for freehand three-dimensional echocardiography." IEEE transactions on medical imaging **21**(9): 1077-1090.

SPSS-Inc. (2008). SPSS Statistics 17.0.

St. Luke's Cataract & Laser Institute. (2008). "Macular Degeneration - Wet and Dry - Age Related Macular Degeneration",Retrieved 30-07-2009, *from* http://www.stlukeseye.com/Conditions/MacularDegeneration.asp.

Staal, J., M. Abràmoff, et al. (2004). "Ridge-based vessel segmentation in color images of the retina." IEEE transactions on medical imaging **23**(4): 501-509.

STARE. (2000). "STructured Analysis of the Retina ",Retrieved 10-3-2010, *from* http://www.ces.clemson.edu/~ahoover/stare/.

Tan, J., P. Mitchell, et al. (2007). "Cardiovascular Risk Factors and the Long-term Incidence of Age-Related Macular Degeneration:: The Blue Mountains Eye Study." Ophthalmology **114**(6): 1143-1150.

Thdibaoui, A., A. Rajn, et al. (2000). A fuzzy logic approach to drusen detection in retinal angiographic images. *in proceedings of* 15th International Conference on Pattern Recognition, Barcelona, Spain.

Tobin, K. W., E. Chaum, et al. (2007). "Detection of Anatomic Structures in Human Retinal Imagery." Medical Imaging, IEEE Transactions on **26**(12): 1729-1739.

Uninova. (2005). "MD3RI - Manual Drusen Deposits Detection on Retina Images", 2010, *from* http://www.uninova.pt/~drusens/md3ri.

Varma, R., S. Fraser-Bell, et al. (2004). "Prevalence of age-related macular degeneration in Latinos: the Los Angeles Latino eye study." Ophthalmology **111**(7): 1288-1297.

Vincent, L. and P. Soille (1991). "Watersheds in Digital Spaces: An efficient Algorithm Based on Immersion Simulations." IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE **13**(6): 583-598.

Vingerling, J. R., A. Hofman, et al. (1996). "Age-related macular degeneration and smoking. The Rotterdam Study." Arch Ophthalmol **114**(10): 1193-1196.

Wall, M. (1996). A genetic algorithm for resource-constrained scheduling. PhD Thesis, MIT

Wang, J., P. Mitchell, et al. (2007). "Epidemiology of Age-Related Macular Degeneration Early in the 21st Century." Retinal Degenerations: 23-59.

Wang, J. J., E. Rochtchina, et al. (2007). "Ten-year incidence and progression of age-related maculopathy: the blue Mountains Eye Study." Ophthalmology **114**(1): 92-98.

Wang, Q., R. Chappell, et al. (1996). "Pattern of age-related maculopathy in the macular area. The Beaver Dam Eye Study." Investigative ophthalmology & visual science **37**(11): 2234.

Wang, Y., J. Nelson, et al. (2003). "Optimal wavelength for ultrahigh-resolution optical coherence tomography." Opt Express **11**(12): 1411-1417.

William, H. P., T. V. William, et al. (2002). Numerical Recipes in C++: the art of scientific computing, Cambridge University Press.

World Health Organization (2007). Visio 2020 The Right to Sight - Global Initiative for the Elimination of Avoidable Blindness : action plan 2006-2011. WHO Press.

Xu, L., Y. B. Li, et al. (2008). "Age-related macular degeneration and mortality: the Beijing eye study." Ophthalmologica **222**(6): 378-379.

Zebulum, R., M. Pacheco, et al. (2001). Evolutionary electronics: automatic design of electronic circuits and systems by genetic algorithms, CRC.

Zhang, C. and A. Wong (1997). "A genetic algorithm for multiple molecular sequence alignment." Bioinformatics **13**(6): 565.

Zhang, X. and G. Fan (2006). Retinal Spot Lesion Detection Using Adaptive Multiscale Morphological Processing. Advances in Visual Computing**:** 490-501.