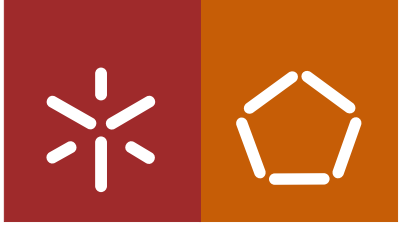
Universidade do Minho
Escola de Engenharia

Jorge Miguel Gomes de Oliveira

Automatic detection of drusen associated with age-related macular degeneration in optical coherence tomography: a graph-based approach

January 2018

UMinho | 2018 Jorge Miguel Gomes de Oliveira Automatic detection of drusen associated with age-related macular degeneration in optical coherence tomography: a graph-based approach



Universidade do Minho
Escola de Engenharia

Jorge Miguel Gomes de Oliveira

Automatic detection of drusen associated with age-related macular degeneration in optical coherence tomography: a graph-based approach

Tese de Doutoramento em Líderes para Indústrias Tecnológicas

Trabalho efetuado sob a orientação do
Professor Doutor Carlos Alberto Batista Silva
e do
Professor Doutor Manuel João Oliveira Ferreira

January 2018

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.
I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, _____

Full name:
_____

Signature:
_____

Acknowlegements

This work was supported by FCT with the reference project UID/EEA/04436/2013, by FEDER funds through the COMPETE 2020 – Programa Operacional Competitividade e Internacionalização (POCI) with the reference project POCI-01-0145-FEDER-006941.

Furthermore, the Portuguese funding institution Fundação Calouste Gulbenkian has conceded me a Ph.D. grant for this work. For that, I wish to acknowledge this institution. Additionally, I want to thank one of its members, Teresa Burnay, for all her assistance with issues related with the grant, for believing that my work was worth supporting and for encouraging me to apply for the grant.

I also want to acknowledge ENERMETER - Sistemas de Medição, Lda and my co-advisor Manuel Ferreira for their support.

I am also grateful to Oftalmocenter and more concretely to Dr. Luís Gonçalves, who was always available for helping me with his medical expertise.

This was work was only possible with the help of my advisor, Carlos Silva, to who I express my sincere gratitude for his guidance, advice and assistance in revising this document.

The moral support of my friends was also crucial for accomplishing this work. I wish to thank in particular Nelson Martins, who has helped me with the revision of this document.

A big "Thank you!" is also due to my family for helping and supporting me throughout my whole life, specially to my siblings, José and Joana. Also want to thank the new members of the family, my niece Sophia and my nephew Rafael, for the joy that they bring me and my family.

At last, I am also grateful to my girlfriend, Mariana, for all the love, affection and emotional support that she has provided me all these years.

This work is not just mine, it also belongs to everyone that was mentioned here. **Thank you very much!**

Abstract

The age-related macular degeneration (AMD) starts to manifest itself with the appearance of drusen. Progressively, the drusen increase in size and in number without causing alterations to vision. Nonetheless, their quantification is important because it correlates with the evolution of the disease to an advanced stage, which could lead to the loss of central vision. Manual quantification of drusen is impractical, since it is time-consuming and it requires specialized knowledge. Therefore, this work proposes a method for quantifying drusen automatically

In this work, it is proposed a method for segmenting boundaries limiting drusen and another method for locating them through classification. The segmentation method is based on a multiple surface framework that is adapted for segmenting the limiting boundaries of drusen: the inner boundary of the retinal pigment epithelium + drusen complex (IRPEDC) and the Bruch's membrane (BM). Several segmentation methods have been considerably successful in segmenting layers of healthy retinas in optical coherence tomography (OCT) images. These methods were successful because they incorporate prior information and regularization. However, these factors have the side-effect of hindering the segmentation in regions of altered morphology that often occur in diseased retinas. The proposed segmentation method takes into account the presence of lesion related with AMD, i.e., drusen and geographic atrophies (GAs). For that, it is proposed a segmentation scheme that excludes prior information and regularization that is only valid for healthy regions. Even with this segmentation scheme, the prior information and regularization can still cause the oversmoothing of some drusen. To address this problem, it is also proposed the integration of local shape priors in the form of a sparse high order potentials (SHOPs) into the multiple surface framework.

Drusen are commonly detected by thresholding the distance among the boundaries that limit drusen. This approach misses drusen or portions of drusen with a height below the threshold. To improve the detection of drusen, Dufour $et\ al.$ [1] proposed a classification method that detects drusen using textural information. In this work, the method of Dufour $et\ al.$ [1] is extended by adding new features and performing multi-label classification, which allow the individual detection of drusen when these occur in clusters. Furthermore, local information is incorporated into the classification by combining the classifier with a hidden Markov model (HMM).

Both the segmentation and detections methods were evaluated in a database of patients with intermediate AMD. The results suggest that both methods frequently perform better than some methods present in the literature. Furthermore, the results of these two methods form drusen delimitations that are closer to expert delimitations than two methods of the literature.

Sumário

A degenerescência macular relacionada com a idade (DMRI) começa a manifestar-se com o aparecimento de drusas. Progressivamente, as drusas aumentam em tamanho e em número sem causar alterações à visão. Porém, a sua quantificação é importante porque está correlacionada com a evolução da doença para um estado avançado, levar à perda de visão central. A quantificação manual de drusas é impraticável, já que é demorada e requer conhecimento especializado. Por isso, neste trabalho é proposto um método para segmentar drusas automaticamente.

Neste trabalho, é proposto um método para segmentar as fronteiras que limitam as drusas e outro método para as localizar através de classificação. O método de segmentação é baseado numa "framework" de múltiplas superfícies que é adaptada para segmentar as fronteiras que limitam as drusas: a fronteira interior do epitélio pigmentar + complexo de drusas e a membrana de Bruch. Vários métodos de segmentação foram consideravelmente bem-sucedidos a segmentar camadas de retinas saudáveis em imagens de tomografia de coerência ótica. Estes métodos foram bem-sucedidos porque incorporaram informação prévia e regularização. Contudo, estes fatores têm como efeito secundário dificultar a segmentação em regiões onde a morfologia da retina está alterada devido a doenças. O método de segmentação proposto toma em consideração a presença de lesões relacionadas com DMRI, .i.e., drusas e atrofia geográficas. Para isso, é proposto um esquema de segmentação que exclui informação prévia e regularização que são válidas apenas em regiões saudáveis da retina. Mesmo com este esquema de segmentação, a informação prévia e a regularização podem causar a suavização excessiva de algumas drusas. Para tentar resolver este problema, também é proposta a integração de informação prévia local sob a forma de potenciais esparsos de ordem elevada na "framework" multi-superfície.

As drusas são usalmente detetadas por "thresholding" da distância entre as fronteiras que limitam as drusas. Esta abordagem falha drusas ou porções de drusas abaixo do "threshold". Para melhorar a deteção de drusas, Dufour $et\ al.$ [1] propuseram um método de classificação que deteta drusas usando informação de texturas. Neste trabalho, o método de Dufour $et\ al.$ [1] é estendido, adicionando novas características e realizando uma classificação com múltiplas classes, o que permite a deteção individual de drusas em aglomerados. Além disso, é incorporada informação local na classificação, combinando o classificador com um modelo oculto de Markov.

Ambos os métodos de segmentação e deteção foram avaliados numa base de dados de pacientes com DMRI intermédia. Os resultados sugerem que ambos os métodos obtêm frequentemente melhores resultados que alguns métodos descritos na literatura. Para além disso, os resultados destes dois métodos formam delimitações de drusas que estão mais próximas das delimitações dos especialistas que dois métodos da literatura.

Contents

List of Figures

List of Tables

List of acronyms

AAD	Absolute Area Difference
AAM	Active Appearance Model
ANN	Artificial Neural Network
AMD	Age-related Macular Degeneration
AREDS	Age-Related Eye Disease Study
AUC	Area Under the ROC Curve
BM	Bruch's Membrane
B-K (algorithm)	Boykov-Kolmogorov (algorithm)
CAD	Computer Aided Detection
CNV	Choroidal Neovascularization
CRF	Conditional Random Field
DMRI	Degenerescência Macular Relacionada com a Idade
DOPU	Degree of Polarization Uniformity
ELM	External Limiting Membrane
FD-OCT	Fourier Domain Optical Coherence Tomography
FT	Fourier Transform
GA	Geographic Atrophy
GCL	Ganglion Cell Layer
GMM	Gaussian Mixture Model
HD	Hamming Distance
ILM	Internal Limiting Membrane
IPL-INL	boundary between the Inner Plexiform Layer and the Inner Nuclear Layer
IRPE	Inner boundary of the Retinal Pigment Epithelium
IRPEDC	Inner boundary of the Retinal Pigment Epithelium + Drusen Complex
IS	Inner Segment

MGDM	Multiple-Object Geometric Model
MRF	Markov Random Fields
MSF	Multiple Surface Framework
MSS	Multiple Surface Segmentation
OCT	Optical Coherence Tomography
OPL	Outter Plexiform Layer
ORNFL	Outer boundary of the Retinal Nerve Fiber Layer
ORPE	Outer boundary of the Retinal Pigment Epithelium
OS	Outter Segment
PCA	Principal Component Analysis
PDT	Photodynamic Therapy
PED	Pigment Epithelium Detachment
PL	Photoreceptor Layer
PPV	Positive Predictive Value
PS-OCT	Polarization-Sensitive Optical Coherence Tomography
QPBO	Quadratic Pseudo-Boolean Optimization
RF	Random Forest
RNFL	Retinal Nerve Fiber Layer
RPE	Retinal Pigment Epithelium
ROC	Receiver Operating Characteristic
SD-OCT	Spectral Domain Optical Coherence Tomography
SHOP	Sparse High Order Potential
SLO	Scanning Laser Ophthalmoscopy
SS-OCT	Swept Source Optical Coherence Tomography
TD-OCT	Time Domain Optical Coherence Tomography
VEGF	Vascular Endothelial Growth Factor
WC	Wavelet Coefficient

Chapter 1

Introduction

In the first section of this chapter, it is substantiated the need for an automatic algorithm of drusen quantification. For that purpose, first it is given a brief description of the eye anatomy and information regarding AMD. Afterwards, it is exposed the relation between this disease and drusen, as well as describing the main techniques for visualizing these lesions. In the second section, the objectives of this dissertation are stated. The last section performs an overview of the following chapters of this document.

1.1 Motivation

The retina is a tissue located at the back of the eye (Fig. 1.1) composed by cells that are capable of transducing light into electrical signals. These signals are sent to the brain, where later are interpreted as sight [7].

Several diseases may affect this tissue, one of the most relevant is the AMD. This is a degenerative[1] disease that affects the macula (Fig. 1.1) [8]. The main risk factor of this disease is aging [9]. AMD can lead to the loss of central vision, which is responsible for vision's fine detail [9]. As a result, the patient can lose its independence and even suffer from depression [9].

In 2010, AMD was the third leading cause of visual impairment worldwide[2] affecting 33 million people [10]. Furthermore, it was the leading cause of visual impairment for the most developed countries in America and Europe with about 4 and 7 million people affected, respectively [10]. This is explained by the higher life expectancy in these countries [9]. In Portugal, it is estimated that about 300 thousand people are affected by AMD [11]. Worldwide, the life expectancy is increasing and thus, it is believed that the number of cases of AMD will also increase; for the period of 2005 to 2030, it is expected a 3 fold increase of AMD cases in patients with 60 year or more [9]. AMD is also a heavy economy burden; in 2010, the overall costs associated with this disease (e.g. loss of workers, medical assistance) were over 250 billion dollars in the whole world [10]. For 2020, it is estimated an increase of 15 % of this cost [10].

[1] A degenerative disease implies that the affected tissue or organ will continuously deteriorate.

[2] The first and second leading cause are cataract and glaucoma, respectively [10].

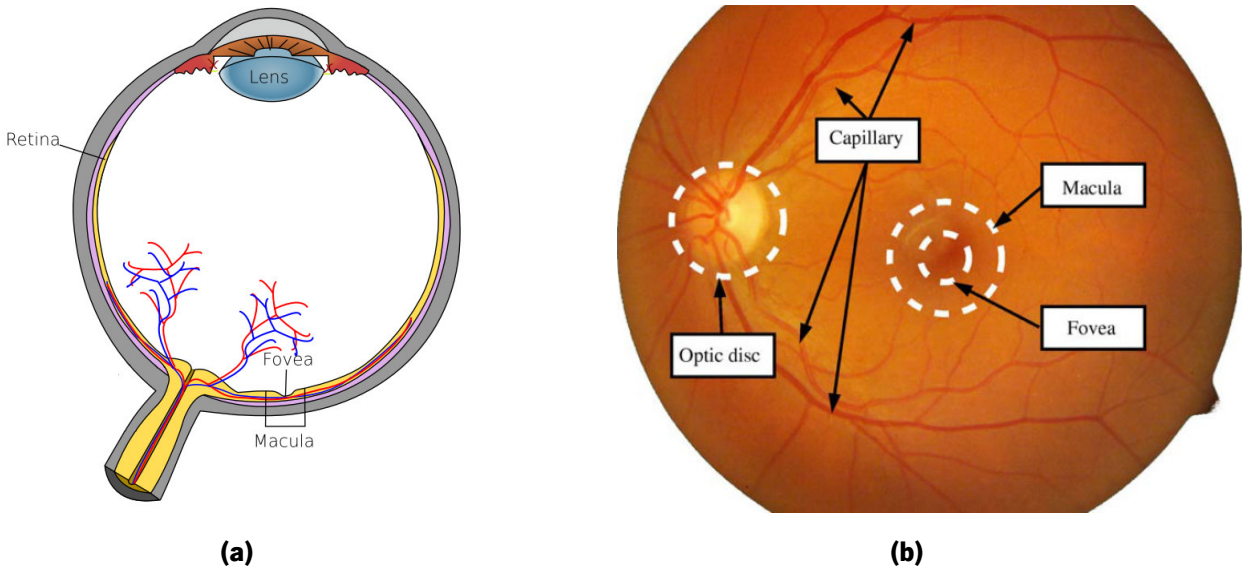


Figure 1.1: A representation of the (a) eye anatomy and a (b) retinography (images adapted from [7]).

The first manifestation of AMD is the appearance of drusen [12]. These structures are accumulations of materials (e.g. proteins and lipids) at the bottom of the retina [13, 14]. Drusen by themselves do not cause vision loss, but it has been established that their increase in size and number is correlated with the progress of AMD [12]. Thus, by evaluating drusen, the physicians can evaluate the progress stage of the patient and adequately manage their condition.

To visualize drusen, the best imaging techniques are digital retinography (Fig. 1.1 (b)) and OCT[3]. The first is a digital photography of the retina, while the second acquires cross sectional images of the retina by measuring the magnitude and time-of-flight of light echoes (similar to ecography). Digital retinography is the current gold standard to visualize drusen [15]. Increasing drusen area in retinography has been associated with an increase in risk of AMD progressing to an advanced stage [16]. Despite being the current gold standard, retinography has some limitations, some of which can be overcome by OCT [15]. For instance, OCT volumes present a higher number of visible drusen and their contours are more defined [15, 17]. Thus, it is thought that the quantification of drusen may be more accurate in OCT volumes than in retinographies [15]. Additionally, OCT acquires cross-sectional images of the retinas that allow the calculation drusen volumes and provide information of both drusen composition and damage of the overlying layers [18].

Drusen quantification can be used for evaluating the progress of AMD. However, manual segmentation of these structures is time consuming for physicians, independently of the imaging technique. Furthermore, the agreement between graders is only moderate [17]. Therefore, the current common clinical practice for drusen evaluation is by visual inspection of retinographies. This is a subjective process and changes may pass unnoticed in follow-up exams. To achieve a quantitative evaluation of drusen in a practical way, several drusen quantification algorithms for retinography have been proposed in the literature [19]. As for OCT, some algorithms

[3]An example of an OCT image can be observed in Fig. 2.7

have also been developed for this purpose[4]. These algorithms can serve as the core component of computer aided detection (CAD) systems that can be used in clinical practice.

1.2 Objectives

The main objective of this work is the development of novel automatic algorithm for quantifying drusen in OCT datasets of patients with intermediate AMD. This algorithm could posteriorly be integrated into a CAD system. Such system would provide physicians with automated drusen quantification, possibly more sensitive to variations than drusen quantification algorithms for retinography [15].

The quantification of drusen can be divided into two objectives: the segmentation of the limiting boundaries of drusen and the detection of these lesions. The segmentation of retinal boundaries in images of AMD is not trivial due to the presence of noise and lesions such as drusen and GA. Thus, the aim is to develop a segmentation algorithm that is robust to these conditions. After segmenting the limiting boundaries of drusen, these lesions have to be located in order to be quantified. Drusen can be detected by thresholding the distance between their limiting boundaries. However, this method excludes small drusen and it is very sensitive to errors in the segmentation. Therefore, the goal is to develop a method that overcomes these issues and detects correctly the vast majority of drusen.

1.3 Thesis overview

The next chapter provides background information to improve the definition of the problem of drusen quantification. In particular, it presents additional information regarding the disease focused in this work (i.e., AMD) and the imaging technique used to visualize the retina (i.e., OCT). Chapter 3 reviews the methods of the literature that were developed for segmenting retinal boundaries in healthy retinas and methods that quantify drusen in images of patients with AMD. Chapter 4 exposes the theoretical background necessary to understand the proposed segmentation and detection methods. Here, the focus is on the framework used for segmenting retinal boundaries and the classifier used for detecting drusen. Chapter 5 describes and evaluates the proposal for segmenting the limiting boundaries of drusen. Afterwards, chapter 6 exposes and evaluates the method proposed for detecting drusen. At last, the concluding remarks are presented in chapter 7.

[4] In chapter 3, a review of drusen quantification methods will be presented.

Chapter 2

Background and Significance

In this chapter, the problem of drusen quantification will be properly framed. First, the anatomy and physiology of the eye is described. Afterwards, AMD is addressed in some detail to understand its origin, consequences, evolution and treatments. In the last section, the main imaging techniques for visualizing drusen will be exposed, with a special focus in OCT.

2.1 Eye anatomy

In the eye, light travels through the cornea, aqueous humour, pupil, lens, vitreous humor and it is finally focused on the retina (Fig. 2.1) [20]. When light reaches the eye, it is considerably refracted by the cornea, a convex lens responsible for about two thirds of the total refraction in the eye [21]. Between the cornea and the iris there is a transparent liquid designated as aqueous humor[1]. Light passes through this liquid practically without any interference [20, 21]. Then, light crosses the pupil which is an opening in the iris [21]. The iris controls the quantity of light gathered by controlling the diameter of the pupil. Thereafter, the light passes through the lens,

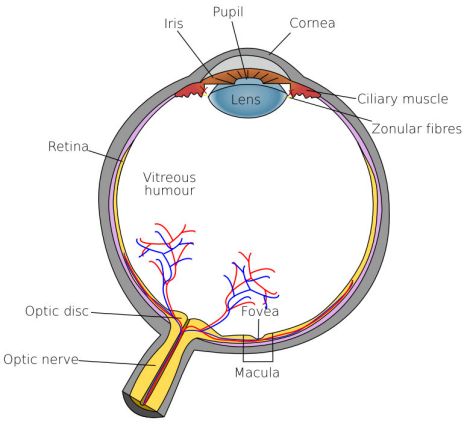


Figure 2.1: A representation of the eye anatomy (image retrieved from [7]).

[1]The aqueous humor serves to keep the cornea firm and slightly curved [20].

whose main function is to correctly focus light in the retina [7, 20, 21]. The shape of the lens is controlled by the ciliary muscle through the zonular fibers (Fig. 2.1). Thus, the lens thickness is increased or decreased if it is necessary to focus objects at short and long distance, respectively[2] [7]. Lastly, light crosses the vitreous humor[3] and reaches the retina. In the retina, there are photoreceptor cells that convert incoming light into electric signals. These signals are posteriorly sent through the optic nerve to the brain, where eyesight is formed.

2.1.1 Retina anatomy

The retina is a tissue organized in several layers and membranes that contains different types of cells (Fig. 2.2). Photoreceptors are the first cells stimulated by light. These are located in the photoreceptor layer (PL), which is one of the outer layers of the retina (Fig. 2.2). Nevertheless, this layer receives almost all the light that reaches the retina, since the anterior layers are transparent [7]. When the photoreceptors receive light, they convert it into an electric signal. This signal is sent to the bipolar cells that relay the signal to the ganglion cells. The ganglion cells deliver the signal to the brain through its axons (Fig. 2.2). These axons compose the retinal nerve fiber layer (RNFL) – a network of more than 1 million axons that bundle into the optic nerve [20].

The structure of the retina is not homogeneous throughout its whole extension; the retina structure is different in areas such as the optic disc, macula and fovea. The optic disc is the most superficial zone of the optic nerve (Fig. 2.1) [7]. This is a blind spot where normal retinal layers do not exist [7]. The macula[4] is an area located in the center of the retina, which is responsible for central vision [7] (Fig. 2.1). This region contains uneven concentrations of different types of photoreceptor cells. In the center of the macula, or fovea, there is a higher concentration of cone cells, which are the cells responsible for the detection of colors [7, 20], while outside the fovea, there is a higher concentration of rod cells, which are very sensitive to light but incapable of discriminating colors [7, 20]. Furthermore, the retina exhibits a depression on the fovea because some layers do not exist in this region, such as RFNL. This depression maximizes the light reaching the photoreceptor cells [20].

The supply of nutrients and oxygen to the retina is performed by the retinal vasculature (around 35 %) and by the choroid (about 65%) [23]. The retinal vessels are mainly found in the inner layers, such as ganglion cell layer (GCL) and RFNL (Fig. 2.2) [24]. These vessels mainly supply the superficial retinal layers, while the outer layers are essentially supplied by the choroid [24], which is a vascular layer adjacent to the retina (Fig. 2.2). The nutrients and oxygen from the choroid have to be transported through the BM[5] and the retinal pigment epithelium (RPE), as depicted in Fig. 2.3. The RPE is the outermost layer of the retina and it is composed by monolayer of tightly bounded pigment cells[6] [9]. The RPE is an important retinal layer, since it is involved in the transportation of oxygen and nutrients from the choroid to the outer layers, absorbs stray light and serves as retinal–blood barrier [9].

[2] This process is designated as accommodation [7, 21]

[3] The main function of the vitreous humor is to support the shape of the eye [21].

[4] The macula has a diameter of 5 to 6 mm.

[5] BM is the innermost membrane of the choroid.

[6] The pigment cells are responsible for the pigmentation of each retina [20]

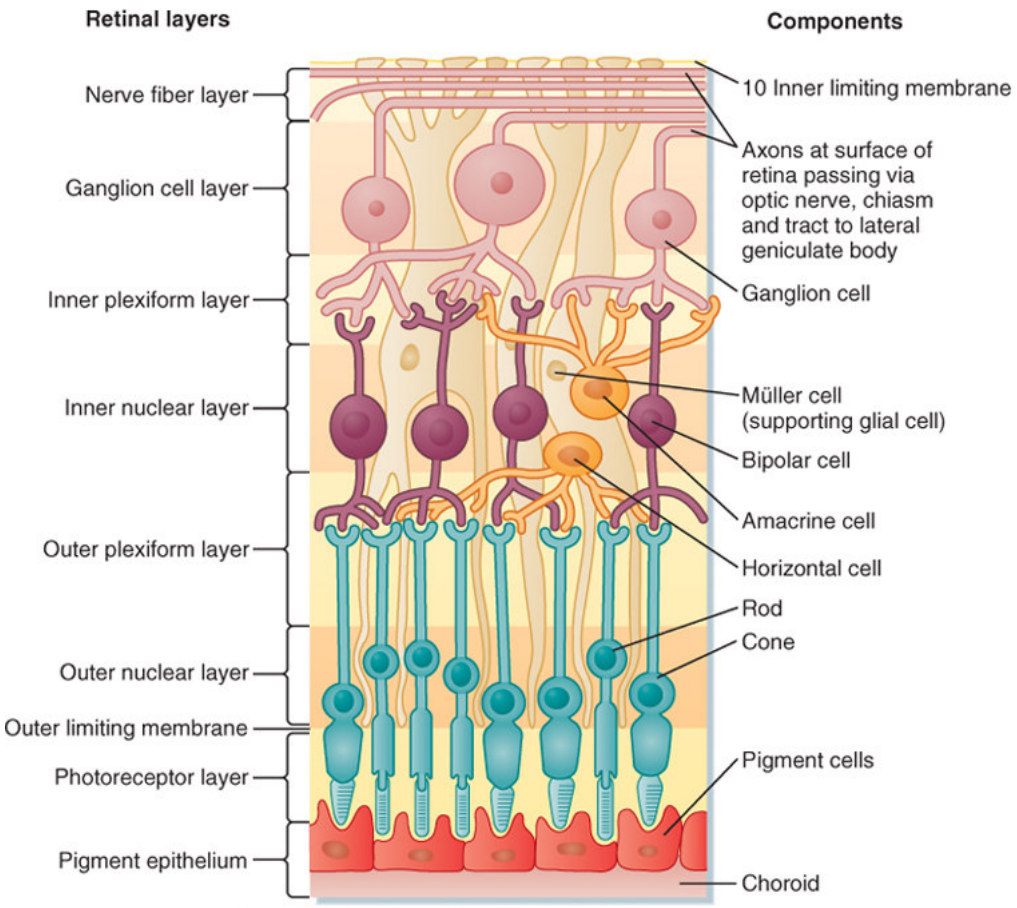


Figure 2.2: Retinal layers and respective cell types (image adapted from [22]).

2.2 Age-related macular degeneration

AMD is a disease that causes degeneration of the macula (Fig. 2.1). AMD's ultimate outcome is the loss of central vision (Fig. 2.4) [9]. In this scenario, the patient cannot distinguish fine detail, which is a severe incapacitation. For instance, without central vision a patient cannot recognize faces, read, write or drive [9]. Even so, AMD does not lead to complete blindness, since the patient still retains peripheral vision (Fig. 2.4).

The causes of AMD are still not fully understood, although there are some pieces of evidence that point to a combination of factors. Some examples of these factors are: oxidative stress, chronic inflammation, metabolic insufficiency and choroidal vascular changes [9].

The most relevant risk factor of AMD is advanced age [12]. However, other risk factors have been associated with AMD, such as: smoking, genetic predisposition, white race, light exposure, obesity, hypertension, high comsuption of vegetable fats and low intake of antioxidants [9, 12].

2.2.1 AMD types and evolution

The progression of AMD can be evaluated through the identification and quantification of some clinical features. Drusen are the first of those features [12]. These are extracellular deposits of material (e.g. lipids and proteins)

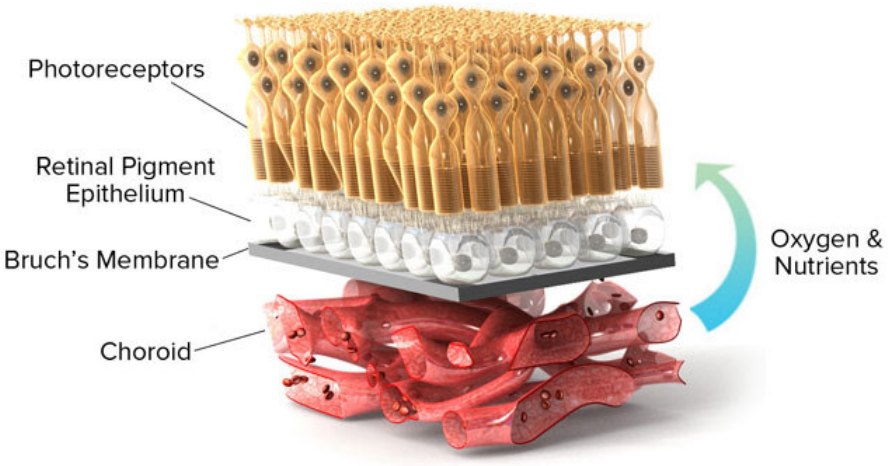
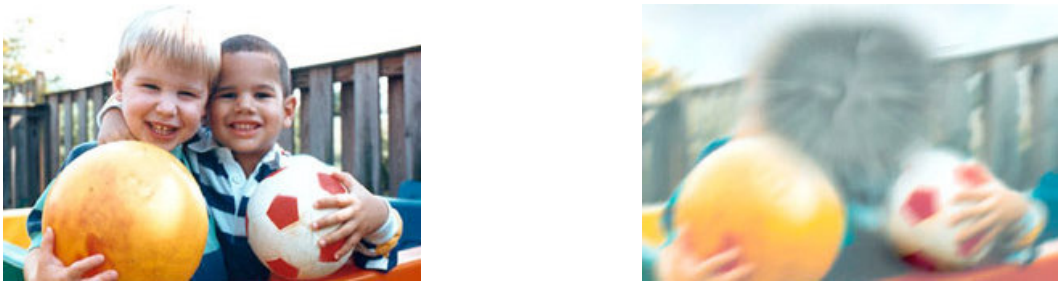


Figure 2.3: Nutrient and oxygen supply from the choroid to the outer retinal layers (image retrieved from [25]).



(a) **(b)**

Figure 2.4: Comparison of vision of a (a) healthy patient with a (b) patient without central vision, the ultimate consequence of AMD (images retrieved from [26]).

that accumulate between the RPE and the BM (Fig. 2.5) [13, 14]. Drusen alone do not cause visual impairment, yet there is evidence that they hinder the function of RPE cells [13]. Cells near drusen frequently present signs of imminent death [13]. Drusen can have quite different shape, size, composition, color, and border definition [9]. These structures can be broadly categorized into hard and soft [9]. Hard drusen are characterized by its reduced size and well defined borders. A few of these drusen can be considered as a normal consequence of aging; 90% of white population over 40 years old is thought to have a least one hard drusen [9]. Soft drusen are usually larger than hard drusen and have indistinct borders [12]. Unlike the previous type, the presence of soft drusen is indicative of AMD [9].

The progression of AMD is categorized in 4 levels according to the age-related eye disease study (AREDS) grading protocol (Table 2.1) [27]. Drusen tend to increase in size and number as the level of the disease progresses (Table 2.1) [12]. Additionally, it has been reported that the total area or volume of drusen is correlated with the risk of progression of AMD to an advanced stage [16, 28]. Therefore, physicians can monitor the progression of AMD by monitoring drusen and some other clinical features. The other relevant features include: pigment abnormalities, pigmented epithelium detachment (PED) and GA (Table 2.1).

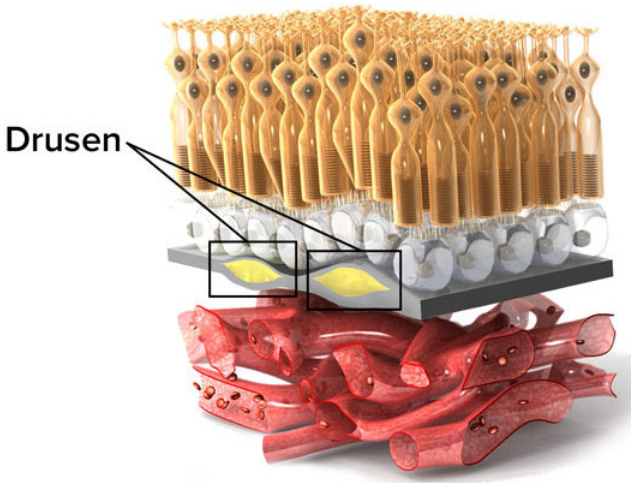


Figure 2.5: Accumulation of extracellular material between RPE and BM cause the formation of drusen (image retrieved from [25]).

Table 2.1: AMD level according to AREDS grading (summarized version). To belong to categories 2 to 4 it is only necessary to have one of the respective clinical features in one eye. Small drusen have a diameter <63μm, intermediate drusen have a diameter between 63 μm and 125 μm and large drusen have a diameter >125μm [27, 29].

AMD level	**Clinical features**
Category 1 (healthy patient)	• No drusen • Nonextensive small drusen
Category 2 (early stage)	• Extensive small drusen • Non extensive intermediate • Pigment abnormalities
Category 3 (intermediate stage)	• Extensive intermediate drusen • Large drusen • Noncentral GA
Category 4 (advanced stage)	**Dry AMD** • Central GA **Wet AMD** • Fibrovascolar or serous PED • Serous or hemorrhagic sensory retina detachment • Subretinal pigment epithelium hemorrage • Scars of confluent photocoagolation

AMD presents two types of advanced stages: dry and wet (category 4 of Table 2.1).

In dry AMD, the cells on the RPE continuously degenerate until forming significant holes, designated as GA [13]. The progressive decline of the RPE compromises the transport of nutrients and oxygen to the photoreceptor cells, leading to the death of these cells (Fig. 2.6 (a)). When the GA occurs outside the center of the macula, this feature is indicative of intermediate stage of AMD progression (category 3 in table 2.1). On the other hand,

(a) (b)

Figure 2.6: Advanced stages of AMD. The characterizing feature of dry AMD is (a) GA that corresponds to the degeneration of the RPE and consequent death of the overlying photoreceptors. In wet AMD there are several characterizing features such as: (b) CNV, PED and intraretinal fluid (images adapted from [25, 32]).

if it occurs at the center of the macula, then GA is the characterizing feature of the advanced stage of dry AMD (category 4 in table 2.1). In this case, the lack of photoreceptor cells cause the blurring of central vision [30]. The dry form represents nearly 90% of the cases of AMD [30]. Fortunately, this form has a slow pace of progression (from months to years) [12]. However, AMD can progress from dry to wet form at any point of the disease progression [30].

In wet AMD, vessels from the choroid start to grow abnormally, possibly as a response of the body to correct the lack of nutrients and oxygen [12]. This occurrence is designated as choroidal neovascularization (CNV) and is one of the characterizing features of wet AMD (Table 2.1) [12]. The new vessels rupture the BM and are able to extend to the subretinal space (below the RPE) or even reach the intraretinal space through breaks in the RPE (Fig. 2.6 (b)) [12]. The structure of the walls of the new vessels is rather fragile allowing leakage, which in turn can cause hemorrhages, edema, PED or exudates [9] (Fig. 2.6 (b)). There are 3 types of PED: drunesoid[7], serous and fibrovascular [9]; only the last two are characteristic of wet AMD (Table 2.1). In a PED, the RPE is detached from the BM [9]. In a serous PED, fluid is accumulated between the RPE and the BM, which can derive from the leakage of new vessels of a CNV (it can also occur without CNV) [9, 31]. In fibrovascular PED, there is also fluid below the RPE, yet the detachment is mainly caused by the proliferation of new vessels of a CNV [31]. Oftentimes, in this type of PED, the RPE has breaks that allow the presence of intraretinal fluid [9]. The fluid accumulations present in wet AMD may cause swelling and scarring of retinal tissues, damaging the central vision [30]. Wet AMD causes 80% of the blindness related to AMD, despite representing barely over 10% of the cases [12]. This demonstrates well the risk that the wet form poses. The severity of this form is related with its fast progression pace (days to weeks) [12]. Therefore, it is important to monitor the progress of AMD and, in particular, for the progression of the dry form into the wet form.

[7] Drunesoid PEDs originates from confluent drusen and liquid retention and are frequently associated with dry AMD [31].

2.2.2 Treatments and lifestyle changes

After the diagnose of AMD, physicians may recommend treatments or lifestyle changes that can slow progression of the disease and improve visual acuity in a few cases. For dry AMD, physicians can propose lifestyle modifications (dietary or otherwise) or the implantation of a miniature telescope [9]. As for wet AMD, besides the same lifestyle modifications, there are several treatments options available. In the following sections, it is present a brief overview of treatments and lifestyle modifications available for both forms of AMD.

2.2.2.1 Dry AMD

The AREDS formulation[8] [9] (composed by antioxidants and zinc) was reported to reduce by 25% the risk of progression of AMD to an advanced stage [9]. Nonetheless, this formulation was only effective for patients in an intermediate or advanced stage of AMD[9] [9]. Moreover, these dietary supplements should be prescribed with caution, since some of its constituents might be hazardous for patients with other diseases [12]. A second study designated as AREDS2 was performed to evaluate a modified AREDS formulation [33]. In this study, it was concluded that the new formulation did not further reduce the risk of progression of AMD, but the substitution of beta-carotene by lutein and zeaxanthin in the AREDS formulation showed lower levels of lung cancer incidence in smoker patients [33].

Besides the ingestion of the AREDS formulation, other lifestyle changes are recommended. Smokers are advised to quit smoking, since this is as a risk factor and there are studies suggesting that non-smokers have a slower progression of AMD [12]. Reduction of fat intake is also advisable because it correlates with increased risk of AMD progression to advanced stages [9]. Furthermore, reduction of sun exposure might be beneficial for those who are exposed to high levels of sunlight [9].

Another treatment option is to replace the lens (Fig. 2.1) with a miniature telescope. The telescope improves the visual acuity of patients by focusing light into a broader region, thus including healthy tissues of the macula [34]. This solution can be applied to patients with intermediate or advanced stages of AMD [35]. For the advanced stage, the device may be used for dry and established wet forms [35].

A more recent therapy is the transplantation RPE cells. One of the main problems of this treatment is finding a donor, thus efforts are being made to produce RPE cells from stem cells [12]. Recently, stem cells were successfully used to improve the visual acuity of a few patients [36].

2.2.2.2 Wet AMD

For wet AMD, there are several possible treatments, being the most common the injection of a drug capable of inhibiting the vascular endothelial growth factor (VEGF), a substance associated with neovascularization [12]. These anti-VEGF drugs are injected into the vitreous humor and bind to VEGF, thus deactivating its angiopro-

[8]The AREDS was a large scale study that also proposed a grading scale for AMD that has been increasingly used (Table 2.1).

[9]Patients with early AMD are not recommend to take the AREDS formulation [9].

liferative effect [12]. The anti-VEGF drugs only slow or stop the progression of wet AMD for period of time. Nonetheless, these drugs can improve the visual acuity of some patients [12].

Another possible treatment is the photodynamic therapy (PDT). This treatment stops the leaking of new vessels by destroying them [12]. A light sensitive dye that binds specifically to the new vessels is intravenously administrated. Afterwards, the laser is pointed to the CNV, activating the drug. The drug causes thrombosis of the new vessels, blocking the leakage and inducing their deterioration [12]. PDT does not slow the progression of AMD; it only reduces the leakage associated with new vessels, thus CNV is likely to appear again. Nevertheless, this is a rather safe method with low occurrence of side effects [12].

Laser photocoagulation is an alternative method for destroying the CNV. This method consists in using a laser to produce heat to destroy the CNV [12]. Unlike PDT, this method is not selective, i.e., it damages healthy tissue around the CNV. After the application of the laser, the focused area is scarred and turned into a permanent blind spot [12]. Formerly, laser photocoagulation was the most common treatment, but nowadays it is only used when the CNV is well defined and it is outside the fovea (to avoid central vision loss) [9, 12].

An approach that is being increasingly widespread is the combination of therapies. This approach consists in combining distinct treatments in order to increase effectiveness of the whole treatment [12]. Since AMD is thought to be caused by several factors, this therapy can improve the overall effectiveness by applying individual treatments that act on different levels [12, 37]. Different combinations of anti-VEGF drugs, PDT or steroids (to act against inflammation) are being tested and some already show evidence of being more effective and needing less medical interventions [12, 37].

There is also the possibility of submacular surgery to remove the CNV. Nonetheless, this technique is rarely used because it suffers from a high rate of relapse [9].

Currently, several new treatments are being investigated [12, 37]. For instance, new drugs that inhibit the growth of new vessels, such as the VEGF Trap-Eye, which has longer effect than previous drugs and require fewer medical interventions [37]. Another example is a drug that silences the gene that is associated with VEGF [12].

2.3 Imaging techniques for drusen visualization

The objective of this work is to detect drusen in order to allow physicians to monitor AMD progression, thus only imaging techniques suited for drusen visualization will be addressed.

There are mainly two imaging techniques suited for drusen visualization: digital retinography and OCT. The digital retinography (henceforward simply designated as retinography) is a digital photography taken through an opthalmoscope (Fig. 2.7). Retinography provides an $enface$ view of the retina where drusen appear as yellowish white areas (Fig. 2.7) [12]. This technique is currently the gold standard for visualizing drusen in non-neovascular AMD [15].

More recently, OCT appeared as an alternative to retinography. OCT acquires cross sectional images (i.e., slices) of the retina by emitting light in the near infrared and measuring the properties of backreflected echoes

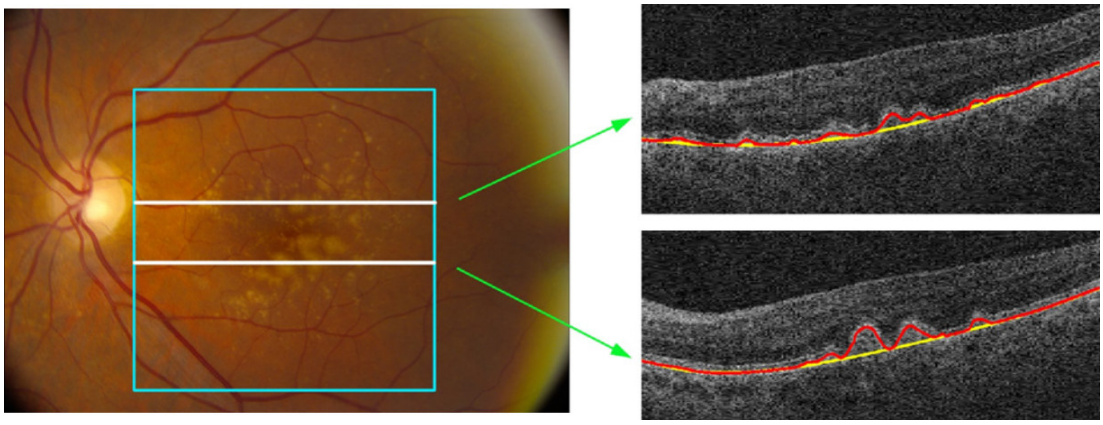


Figure 2.7: Drusen visualization in retinography (left image) and OCT (right image). The OCT slices are represented in the retinography by white lines. The RPE is segmented in the OCT images; the red line represents the anterior boundary and the yellow line the BM. In retinography the drusen are yellowish white lesions and in the OCT images they are elevations of the red line relatively to the yellow line (image adapted from [39]).

(Fig. 2.7) [38]. OCT can acquire several cross sectional images to form a 3D representation of the retina. With this information, drusen volumes can be computed, which is related with the progress of AMD [28]. Furthermore, it seems that in OCT, the number of drusen detected is superior than in retinography [18, 17]. This might be explained by the variations in background in retinography and by the sharper boundaries of soft drusen in OCT images [15]. In addition, OCT can provide information about the material of the drusen (correlated with the magnitude of the echo) and the overlying layers, which are thought to be damaged when there is a drusen nearby [18]. On the other hand, in OCT, it is difficult to detect hard drusen due to its small dimension, which is frequently comparable to the distance between slices [15].

In the next section, it will be presented a more detailed contextualization of OCT and description of its working principle.

2.3.1 Optical coherence tomography

OCT consists in the emission of light and measurement of the time-of-flight and magnitude of the backreflected echoes. In image rendering, the magnitude corresponds to a shade of gray and the time of flight provides information about the location of origin of the echo.

This imaging technique is very similar to ultrasound. Nonetheless, OCT provides a better resolution, i.e., around 1-15 μm (Fig. 2.8) [38]. This resolution allows the visualization of materials' microstructure, which is particularly interesting for medical applications. Due to this capacity, the importance of OCT has been increasing in the health sector. Although its usage is not restricted to health-care, OCT was primarily meant for this sector [38]. In the beginning, the idea of this modality was to perform an $in\ situ$ biopsy, since it allows the visualization of tissue microstructure without excision [38]. OCT can be used when excisional biopsy is not viable, namely

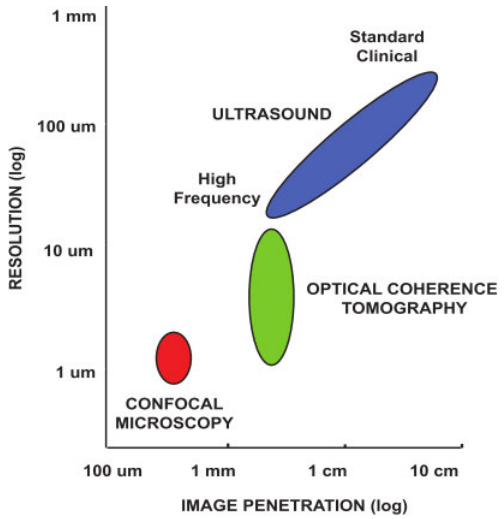


Figure 2.8: Resolution and image penetration for the OCT, ultrasound and confocal microcopy (image retrieved from [38]).

when it is hazardous (e.g. opthalmology) or there is an elevated sampling error[10] (e.g. detection of cancer) [38]. This imaging technique is presently being used to image several anatomic structures, such as: retina, anterior eye, gastrointestinal structures, coronary arteries, skin, teeth, among others [38].

The OCT emergence filled a gap between ultrasound and confocal microscopy (Fig. 2.8). OCT has a better resolution than ultrasonography, however it possess lower image penetration. When comparing to microscopy, OCT has worse resolution, yet it can penetrate deeper. In sum, there is a trade-off between resolution and imaging depth (Fig. 2.8). Before the appearance of OCT, there was a gap between ultrasonography and confocal microscopy. Nowadays, the OCT introduces an new range to this trade-off that has been proving to be useful in several applications.

2.3.1.1 Working principle

In this section, the main concepts to understand the OCT systems will be presented, as well as different types of OCT systems.

The light used in OCT is usually in the range of the near infrared (wavelengths between 800 nm and 1000 nm). This choice is mainly related with the existence of an optical window for human tissue (Fig. 2.9). In this optical window, water and the main chromophores of the human tissue[11] absorb relatively low quantities of light [40]. In the wavelengths of the optical window, light is less attenuated and therefore, it can penetrate deeper into the tissues.

The choice of infrared wavelengths is also influenced by scattering, which has magnitude of two to three times

[10] Biopsy sampling error occurs when the tissue extraction misses the lesion. OCT can be used to guide the biopsy, in order to reduce this error [38].

[11] The main chromophores of human tissue are hemoglobin and melanin. [40]

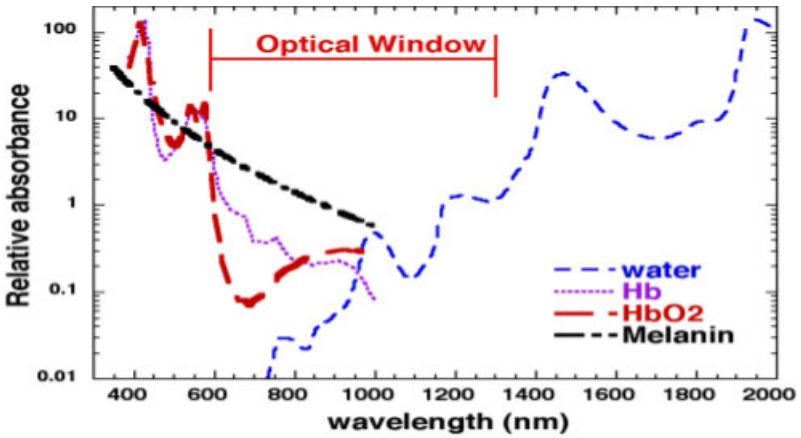


Figure 2.9: Tissue optical window. The infrared light has wavelengths between 800 nm and 1000 nm. For wavelengths lower than 1000 nm the absorption is mainly due to melanin and hemoglobin, while for higher wavelength the absorption is governed mainly by water (image retrieved from [40]).

higher than tissue absorption. The scattering coefficient (μ_s) can be modeled by the following the expression:

$$\mu_s = a\left(\frac{\lambda}{500}\right)^{-b}, \tag{2.1}$$

where λ is the wavelength, a is the attenuation coefficient for 500 nm and b is the scattering power[12] [41]. The Eq. 2.1 indicates that scattering is less pronounced for large wavelengths (Fig. 2.10), allowing light to penetrate deeper into the tissues [38]. However, this has the side effect of deteriorating axial and transverse

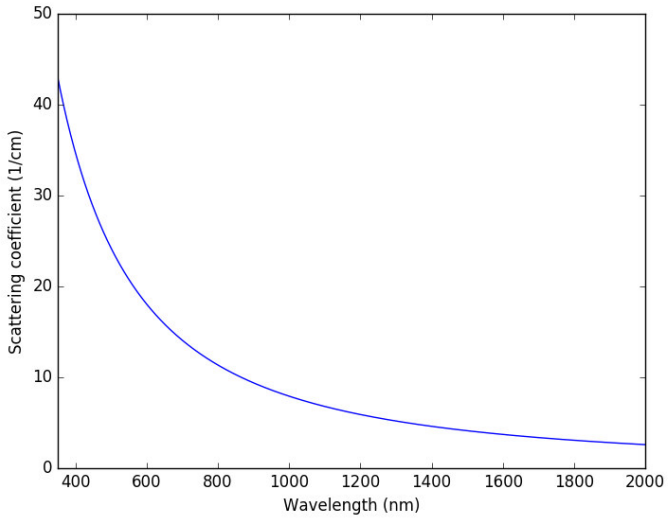


Figure 2.10: Relation between the scattering coefficient and wavelength for brain tissue. The values of a and b for brain tissue were considered, i.e., $a = 24.2$ cm^{-1} and $b = 1.611$ [41]. Brain tissue was selected because the retina is mainly composed of neurons [24].

[12]The parameters a and b were determined experimentally.

resolutions [38]. So, the center wavelength of the light source should take into account these factors, as well as the application. For instance, in retina imaging the light source is normally centered around 830 nm. This wavelength was selected because of water's low absorption[13] (Fig. 2.9) and to have a good resolution for the depth necessary to image the eye [38].

The main difference between the OCT and ultrasonography is the use of light instead ultrasounds. This implies that wave propagation speeds are very uneven. The speed of sound is around 1500 m/s, while in the case of light is around 3×10^8 m/s [38]. In ultrasonography, to obtain a resolution around 0.1 mm, it is necessary a time resolution of 0.1 μs, which is possible with current electronic systems [38]. When using light, for a resolution of 10 μm, the time resolution needs to be 30×10^{-15} s [38]. This time resolution is not possible with present electronic systems and so, it is necessary to utilize optical techniques. In OCT, interferometry is usually used to measure the properties of the echoes. Interferometry is the core technique used in current OCT systems, either in time domain OCT (TD-OCT) or Fourier domain OCT (FD-OCT). In the next subsections these two types of systems will be covered.

2.3.1.2 Time domain OCT

Interferometry is a measuring method based on the interference of waves [42]. OCT uses a Michelson interferometer to superimpose the wave of the reference arm to the wave of the sample arm (Fig. 2.11) [43]. In this interferometer, the light source emits a light beam towards the beam splitter, which splits the beam into two (50% intensity for each one). One beam travels to the sample and part of it is backreflected, forming echoes[14]. The other beam travels in direction of the reference mirror and it is fully reflected. The two waves meet at the beam splitter where the interference occurs.

Interference is only stable when the waves are coherent or partially coherent, i.e., the difference between their phases is constant [45]. If the phase difference between waves is random, then they are non-coherent.

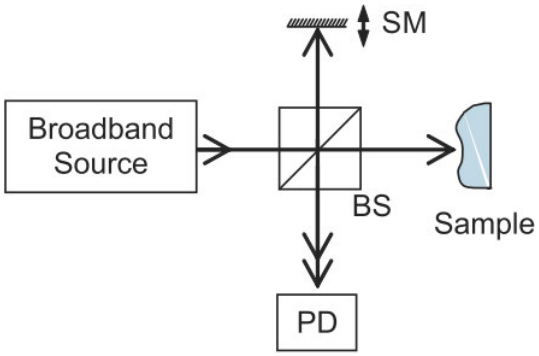


Figure 2.11: Schematic of TD-OCT. From the beamsplitter until the sample is the sample arm and from the beamsplitter to the reference mirror is the reference arm. SM is the scanning mirror, the BS is the beamsplitter and PD is the photodetector (image adapted from [44]).

[13] Light needs to cross the vitreous humor which is mainly water.

[14] The other part is either absorbed or reflected in other directions (due to scattering).

Furthermore, two waves from the same source are only coherent, if the phase distance between them is lower than the coherence length, a quantity that is specific of each light source [43]. This implies that interference only occurs when the difference between the paths of the reference and sample arms are within a coherence length. Therefore, the sample can be scanned for different depths by moving the mirror of the reference arm.

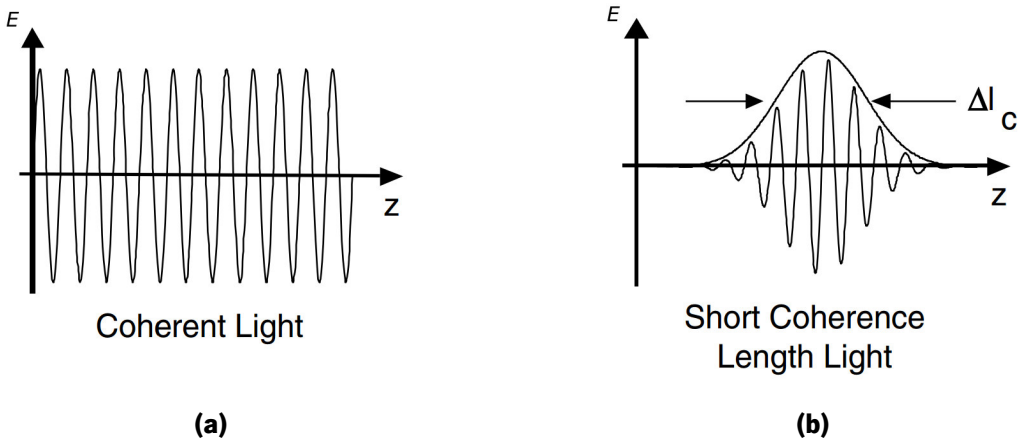


Figure 2.12: Interference result with light sources with (a) long coherence length and with (b) short coherence length. Interference only occurs if the path lengths of the two arms are within one coherence length. This explains why for light sources with long coherence lengths there is interference for all depths, yet it only occurs for some depths when using light sources with short coherence [43]. E stands for electric field, z for sample depth and Δl_c for coherence length (image retrieved from [43]).

After the interference occurrence, the result is registered through a photodetector (Fig. 2.13). The resultant signal is similar to what is presented in Fig. 2.12. Afterwards, the signal is demodulated in terms of amplitude, resulting in an A-scan (Fig. 2.13) [43]. The A-scan is a vector of amplitude values. These values are correlated with the intensity of the echoes and their positions correspond to the locations in the axial dimension where the echoes originated [43]. There is a scanning mechanism that focus the light for other lateral positions, thus several A-scans are acquired to form a B-scan (Fig. 2.13) [43].

In OCT, the axial and transverse resolution are decoupled and thus, they depend on different factors. The axial resolution is equal to the coherence length and it is given by [38]:

$$\Delta z = \frac{2\ln 2}{\pi}\frac{\lambda_0^2}{\Delta\lambda}, \tag{2.2}$$

where λ_0 is the center wavelength of the light source and the $\Delta\lambda$ is the full-width-at-half-maximum of the light source spectrum. The coherence length is correlated with the bandwidth of the beam. Light sources with narrow bandwidths (e.g., lasers) have high coherence lengths, on the other hand light sources with a broader bandwidth (e.g., super-luminescent LED) have lower coherence lengths [46]. The coherence length also depends on the center wavelength; lower center wavelengths imply short coherence lengths [38]. Ideally, a light source with a wide spectrum and low center wavelength should be used to obtain a good axial resolution [44]. The transverse

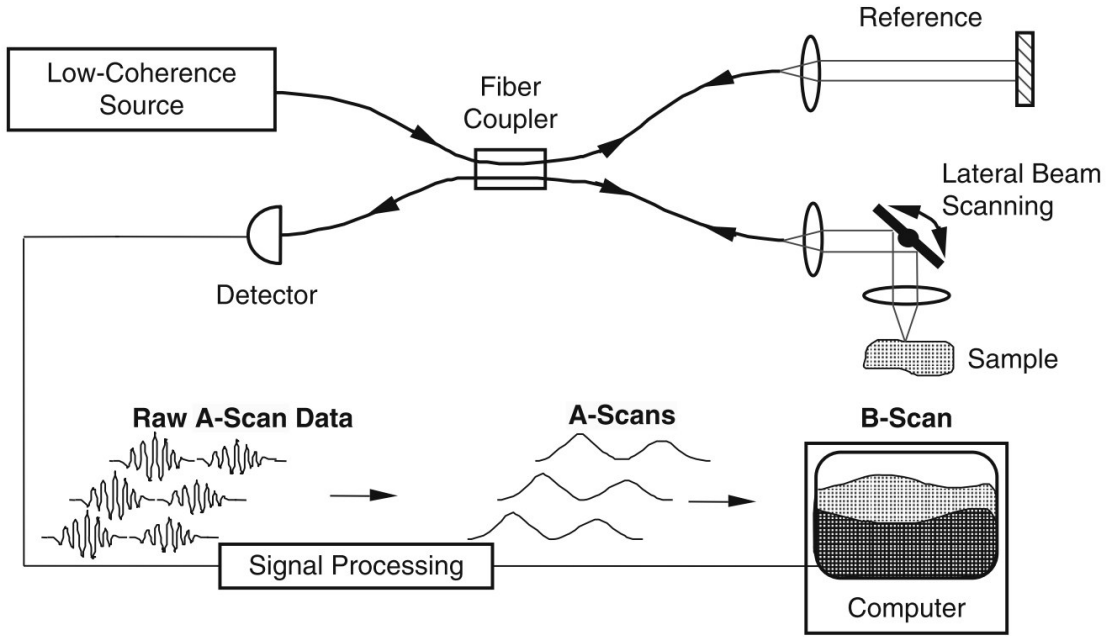


Figure 2.13: Schematic of generic fiber optic OCT. The bold lines represent optic fiber, while the thin lines of the reference and sample arms correspond to free-space optical paths (image retrieved from [43]).

resolution is given by [47]:

$$\Delta x = \frac{2\sqrt{\ln 2}}{\pi}\frac{\lambda_0}{NA}, \tag{2.3}$$

where λ_0 is the center wavelength of the light source and NA is the numerical aperture[15]. Assuming that the wavelength is fixed for a given application, the transverse resolution can be improved by having a focusing lens with high numerical aperture, thus focusing the beam into a small spot of the sample (Fig. 2.14) [47]. However, there is a trade-off between the transverse resolution and the depth of focus[16] (b), which is given by [38]:

$$b = \frac{\pi \Delta x^2}{2\lambda_0}, \tag{2.4}$$

where Δx is the transverse resolution and the λ_0 is the center wavelength of the light source. If the numerical aperture is increased, the transverse resolution is improved, yet the depth of focus is reduced (Fig. 2.14) [47]. Optical coherence microscopy, a variant of OCT, uses lens with high numerical aperture to improve the transverse resolution. This variant is mainly suitable for $en\ face$ imaging, despite having some depth of focus (better than confocal microscopy) [38].

2.3.1.3 Fourier domain OCT

In TD-OCT, the echoes are scanned for defined depths (controlled by the reference mirror), still these originate from the whole sample. Therefore, excessive light is emitted to the sample and the acquisition time is limited by

[15] Numerical aperture is the ability of a lens to gather or focus light. It is given by $NA = n * sin(\alpha)$, where n is the refraction index of the medium between the lens and the sample and α is half of the opening angle of the objective [48].

[16] Depth focus is the image plane range in which the image is not out of focus (for a still object) [49].

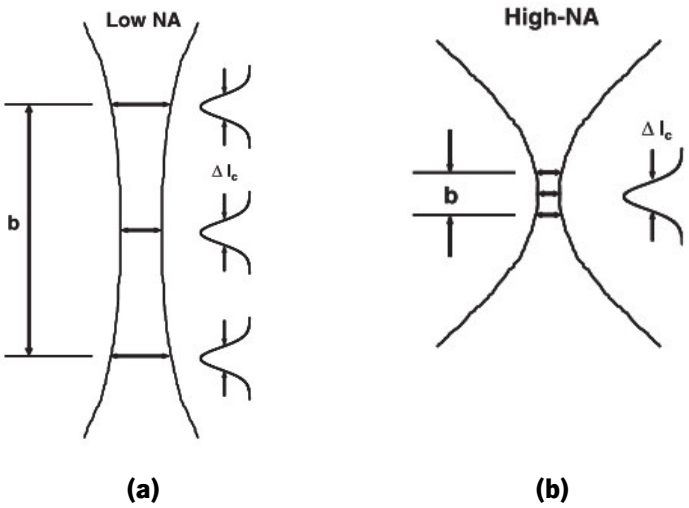


Figure 2.14: Transverse resolution in OCT. In (a) a lens with a low numerical aperture is used, while in (b) one with a high numerical aperture is utilized. The parameter b is the depth of focus. In (a) the transverse resolution is better than in (b), on the other hand, the depth of focus is longer in (b) than in (a). Δl_c stands for coherence length (images adapted from [38]).

scan speed [38]. The FD-OCT tries to overcome this limitations by obtaining the signal in the frequency domain [38]. There are two distinct types of FD-OCT (Fig. 2.15): spectral domain OCT (SD-OCT) and swept source OCT (SS-OCT).

The schematic of the SD-OCT is similar to the TD-OCT (Fig. 2.15). The main differences are the immobilization of the reference mirror and the substitution of the photodetector by a spectrometer (Fig 2.15). The power

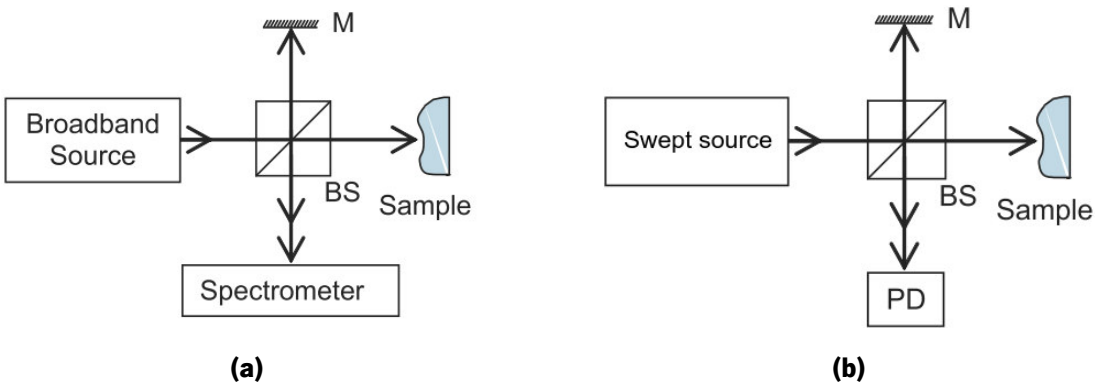


Figure 2.15: Fourier domain OCT. Schematics of (a) SD-OCT and (b) SS-OCT. M is the fixed mirror, BS is the beamsplitter and PD is the phothorecepor (images adapted from [44]).

of the signal detected by the spectrometer for a given wavenumber k[17] is given by[18] [44]:

$$P_{detector}(k) = P_{ref}(k) + P_{sig}(k) + 2\sqrt{P_{ref}(k)P_{sig}(k)}\cos(2k\Delta L), \tag{2.5}$$

where $P_{ref}(k)$ and $P_{sig}(k)$ correspond to the power for the reference and sample arm, respectively. ΔL is the difference in path lengths between the two arms. This equation demonstrates that the power spectrum of signal resultant from interference modulates the difference of path lengths between the two arms in frequency. Therefore, the reference mirror can be fixed because the information of depth is encoded in the frequency components of the power spectrum [44]. The full depth profile, or A-scan, can be retrieved by from the Fourier transform (FT) of power spectrum (Fig. 2.16) [44].

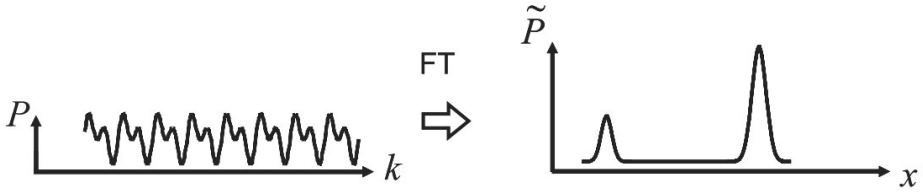


Figure 2.16: Acquiring information of echo properties from the power spectrum of the interfered signal (detected by the spectrometer). The magnitude and the origin location of the echoes are related with the FT of the interfered signal. P stands for power, k for wavenumber, x for sample depth and $\widetilde{P}$ for the magnitude of the FT of the power signal (image adapted from [44]).

By simultaneously detecting echoes from all depths of the sample, the SD-OCT speed is 50 to 100 times faster than the TD-OCT [38]. The higher axial scan rate allows the reduction of motion artifacts in SD-OCT, while rendering feasible a 3D reconstruction of the sample with an exam of a few seconds [38]. Additionally, the simultaneous detection increases the sensitivity of the detector, since the time of integration of the detector can be extended [44]. Another advantage of SD-OCT is the reduced radiation emission, which can be hazardous for the tissues (e.g. retinal tissue).

The SS-OCT is another type of FD-OCT that has resembling capabilities with the SD-OCT. The main advantage of the SS-OCT is the better signal-to-noise ratio that is possible due to the lower presence of noise [44]. The SS-OCT is similar to the SD-OCT, however it uses a tunable narrowband light source (e.g. tunable laser) instead of a broadband light source and a spectrometer in the place of a photodetector (Fig. 2.15) [44]. The tunable source is rapidly swept for a range of wavelengths that are then detected by the photoreceptor. A photodetector is used instead of a spectrometer because the signal is acquired at fixed points in time that possess different wavelengths. These wavelengths are known because the rate of the sweep of the light source is also known. Distinct wavelengths correspond to different depths of the sample [44]. A-scans are obtained similarly to SD-OCT, i.e., by applying the FT to the power spectrum of the detected signal. In this case, the discrete version of the FT is employed, since the signal is acquired for a fixed number of wavelengths [44].

[17] Wavenumber corresponds to $2\pi/\lambda$, where λ is the wavelength.

[18] This expression assumes a top-hat bandwidth [44].

2.4 Summary

This chapter started by introducing the anatomy and physiology of the eye, with a special focus on the retina. This information was presented to contextualize AMD. The causes of this disease are still not fully understood, but its consequences are clearly visible on the macula. Clinical features such as drusen, GA and PED affect the outer layers of the macula. When in an advanced stage, AMD can lead to loss of central vision, a severely incapacitating condition. The advanced stage can have two forms: dry or wet. The dry form has a slow progression, but the disease can evolve to a wet form at any point. The wet form is particularly dangerous because it has a fast progression. Some treatment options are available for both forms, but they only stop or slow the progression of AMD. Therefore, it is particularly important to monitor the development of the disease. AMD can be monitored by detecting and quantifying drusen. This can be performed manually, which is very time consuming, or using automatic algorithms in retinographies or OCT volumes. Retinography is the current gold standard for drusen visualization, however OCT has some advantages over this imaging modality, such as well defined drusen contours and higher sensitivity for detecting drusen. OCT devices can operate in time and Fourier domains. Those that operate in the Fourier domains, such as the SD-OCT, can avoid the use of excessive amounts of light, while reducing the time duration of the exam.

Chapter 3

Literature review

In the previous chapter, it was motivated the need for automatic quantification of drusen. A fundamental task for quantifying drusen material is the segmentation of its limiting boundaries (Fig. 3.1). Several algorithms for segmenting retinal boundaries have been proposed in the literature. Most of them were developed for volumes of healthy retinas. This type of algorithm can be useful for measuring the retinal thickness, as well as the thickness of each retinal layer. The algorithms developed for healthy retinas assume that the boundaries are well defined, smooth and approximately parallel. However, that is not always the case in volumes of diseased retinas (Fig. 3.1) [50]. Therefore, these algorithms frequently exhibit coarse errors in these type of volumes, which are caused by the presence of lesions or other morphological changes related with diseases [50]. Thus, there is a need for algorithms that can cope with clinical features related with retinal diseases. In the literature, there several examples of this type of algorithms. Usually, they are based in methods used for healthy retinas, implying that there is a considerable knowledge overlap between algorithms designed for healthy and diseased retinas.

This chapter reviews the main methods used for the segmentation of boundaries in healthy retinas. Furthermore, it also reviews the algorithms that can segment retinal boundaries in the presence of morphological changes caused by AMD (Fig. 3.1); in particular those that perform drusen quantification.

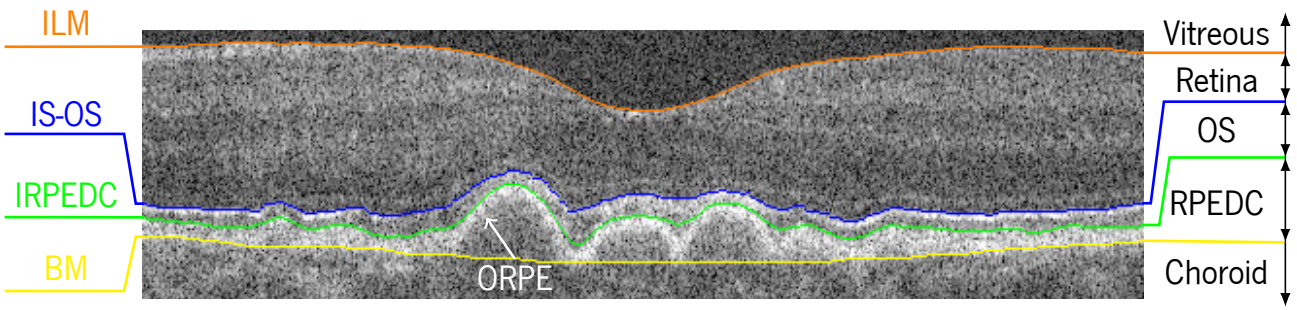


Figure 3.1: Example of an automatic segmentation of an OCT image in the fovea region with the presence of large drusen. IS-OS represents the boundary between the inner segment and the outer segment of the photoreceptor layer, while retina layer refers to the inner layers of the retina.

3.1 Segmentation of boundaries in healthy retinas

The segmentation of boundaries of healthy retinas is one of the most relevant topics in image processing for OCT volumes. Several methods have been proposed in the literature [51, 52], which can be categorized by the types of segmentation techniques. Here, the methods are divided into the following categories: peak detection, active contours, statistical modeling, pixel classification, shortest path and multiple surface framework (MSF). In the following sections, some of the most relevant works of each category will be reviewed.

3.1.1 Peak detection

Peak detection is possibly the earliest and simplest approach for segmenting retinal boundaries. The algorithms within this category segment the boundaries by searching for peaks and valleys in the vertical direction of an image, such as the denoised intensity gradient (Fig. 3.2). Usually, these methods are rather dependent on the pre-processing used, namely the filter used to attenuate the speckle noise. Without denoising, these methods would detect peaks and valleys caused by speckle noise. Furthermore, the peak detection methods tend to be sensitive to image discontinuities, such as the shadowing caused by blood vessels.

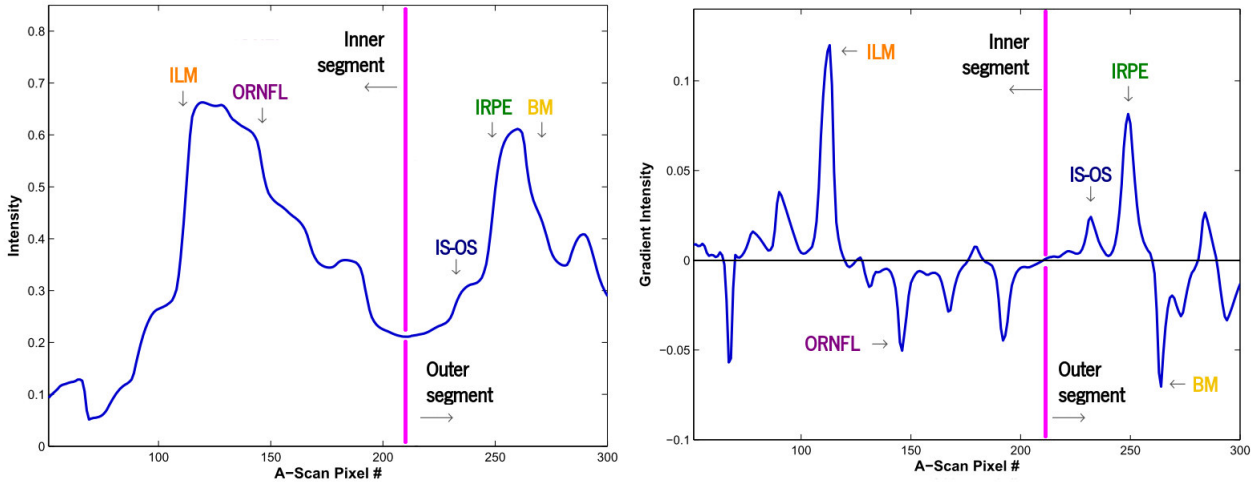


Figure 3.2: Example of the (a) intensity and (b) gradient profiles of a denoised A-scan. The inner and outer segments correspond to the division of the retina that Mayer $et\ al.$ performed in [53], which is not to be mistaken with the inner and outer segments of the PL (images adapted from [53]).

Koozekani $et\ al.$ [54] proposed one of the first methods for the segmentation of retinal boundaries. This method was developed to only segment the ILM and the inner boundary of the RPE (IRPE). Initially, the image is denoised by applying a median filter twice. Then, a 2^{nd} derivate Gaussian kernel is used to enhance edges column-wise. The resultant image is searched for the two most prominent edges of each column. Due to noise and artifacts, the resultant segmentations do not always correspond to the ILM and IRPE. To correct coarse

errors, the segmentations of both boundaries are divided into segments. This division assumes that transitions with 10 pixel or more limit each segment. The segments of coarse errors are corrected using a Markov boundary model, thus the height of each column is based on an weighted average of the neighbors' height derivatives. The segmentations are then fitted to a cubic B-spline, the result of this operations is a couple of smooth segmentations, which are used as final result. This algorithm was later modified by Srinivasan $et\ al.$ [55] to segment several boundaries of the outer retina.

Cabrera $et\ al.$ [56] proposed a peak detection method that segments 7 boundaries. The algorithm starts by removing the DC bias and noise at the background level. Afterwards, a complex diffusion filter is applied to attenuate noise. Then, a coherence enhanced diffusion filter is applied to the denoised image to extract the structure coherence matrix. In this matrix, edges are enhanced according to their prominent local orientation (Fig. 3.3). The information of local orientation can connect interrupted edges, partially avoiding the issue of image discontinuities (e.g., shadows caused by vessels). The segmentation of the boundaries is performed by searching for peaks and valleys on the structure coherence matrix (Fig. 3.3). Even with the information of the structure coherence matrix, the extensive image discontinuities cannot be properly segmented. Therefore, in these regions, a linear interpolation is used to connect the available boundary segments.

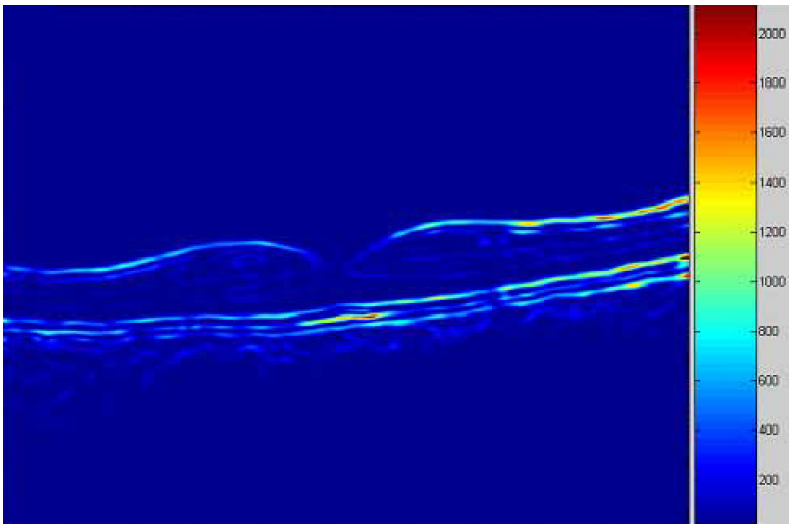


Figure 3.3: Example of the structure coherence matrix determined for the coherence enhanced diffusion filter of a denoised OCT image (image retrieved from [56]).

Mayer $et\ al.$ [53] proposed another peak search method for evaluating the progress of glaucoma. This disease does cause severe alterations to the retinal morphology, nonetheless its progress can be evaluated by measuring the thickness of the RNFL. Thus, the ILM and outer boundary of the RNFL (ORNFL) are segmented. Additionally, the outer boundary of the RPE (ORPE) is also segmented, allowing the calculation of the retinal thickness. Initially, the image is separated into inner and outer segments. The boundary of these regions is found by searching column-wise the minimum intensity located between the largest maximums in an denoised image (Fig. 3.2). Afterwards, the ILM and ORPE are found as the highest peak of the inner segment and the lowest valley of the outer segment, respectively (Fig. 3.2). The ORPE is used to flatten the image. After

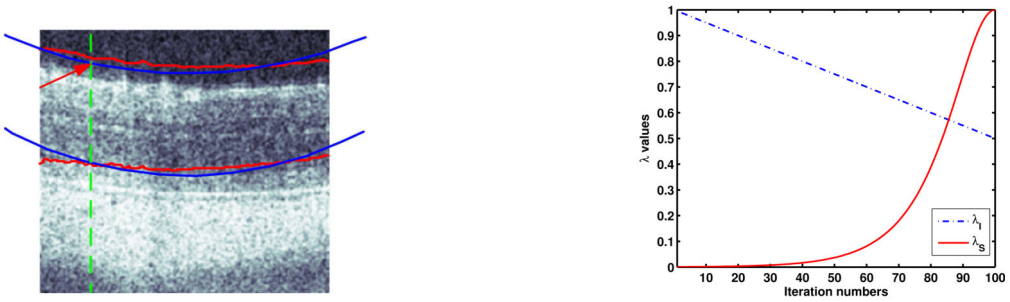
the flattening, the ORNFL is segmented through a minimization procedure that considers the gradient and two smoothness terms. One of those terms avoids the presence of sharp transitions on the segmentation, while the other tries to minimize the differences of the segmentation to its mean value[1]. The segmentation method was altered because the empirical results for the ORNFL using just peak searching were poor; this may be related with the lower contrast of this boundary and the presence of vessels nearby. At the end, all boundaries are smoothed with a scheme based on interpolation and filtering.

Other methods, such as [57, 58], also describe simple algorithms based on peak search used to study features of medical relevance for glaucoma.

3.1.2 Active contours

Active contours algorithms use variational methods to determine a curve that minimizes a certain energy functional. The energy functional must be designed so the determined curve corresponds to the desired boundary. Besides the information of the image, the energy functional may also consider regularization and prior information, such as shape and relative position to other boundaries. This additional information increases the robustness of the method to noise and artifacts.

Yazdanpanah $et\ al.$ [2] proposed an active contours method that does not use information of edges. The proposed method is based on the Chan-Vese method [59]. The simultaneous segmentation of boundaries is performed through the minimization of an energy functional composed by region, shape and regularization terms. The region term encourages each segmentation region to have a homogeneous intensity. The shape term encourages the segmentation to follow a predefined shape, in this case a circular arc (Fig. 3.4 (a)). The smoothness term introduces regularization to the segmentations. The weights of the regions and shape terms are adapted according to the number of iterations – in the initial iterations the weight of the region term is higher than the weight of the shape term (Fig. 3.4 (b)). This is done so the segmentation initially encloses homogeneous



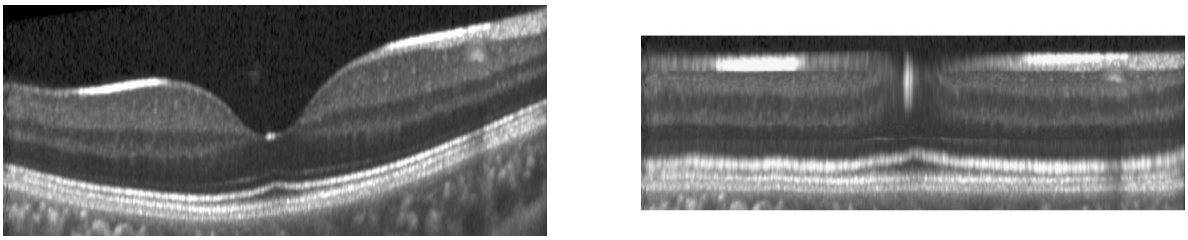
(a) **(b)**

Figure 3.4: Details from the method proposed by Yazdanpanah $et\ al.$ [2]. The image (a) shows the segmentation of two retinal boundaries in red and the circular arc used for the shape prior in blue. The image (b) exhibits the adaptation of weights of the region term (λ_I) and the shape term (λ_S) according to the number of iterations (images adapted from [2]).

[1]With this, the author assumes that the ORNFL is approximately flat after the flattening operation.

regions and, in the last iterations, it approaches a circular arc (Fig. 3.4 (a)), which increasing robustness to noise and artifacts. The weights of the shape term are also spatially adapted by the gradient magnitude – regions with lower gradients have larger shape terms. This adaption increases the influence of the shape term when the information of the image is poor.

Carass $et\ al.$ [60] adapt the multiple-object geometric model (MGDM) [61] to the segmentation of retinal boundaries. The algorithm starts with a coarse segmentation of the ILM and the BM by peak search in a smoothed image. With the segmentation of these boundaries, they convert the volume to a flat space. In this space, all retinal boundaries are approximately flat, even in the fovea region (Fig. 3.5). Since the the algorithm operates in the flat space, the initial estimations of the boundaries are just evenly spaced flat surfaces. Then, the segmentation is performed simultaneously for all boundaries using the MGDM. The MGDM contains a constraint based on digital homeomorphism that maintains the relations between segmentations. In this case, that implies that ordering of boundaries is maintained. The MGDM uses information of the gradient vector flow of the probabilities obtained by a random forest (RF) classifier. The classifier is based on the work of Lang $et\ al.$ [62] (presented in section 3.1.4), in which the classifier outputs probabilities of a pixel belonging to a boundary. The MGDM also has a term for regularization based on the curvature of the segmentations, which encourages them to be flat.



(a) **(b)**

Figure 3.5: Conversion of (a) OCT B-scan to the (b) flat space (imageretrieved from [60]).

Novosel $et\ al.$ [63] exposed another approach for using active contours. In this work, the authors start by compensating the raw OCT data for noise and depth decay. Then, the image is converted to attenuation coefficients[2]. In this work, they also detect the optic nerve head and the location of blood vessels, so the results are only evaluated outside these structures. The initial boundaries for the active contours method are defined using the Dijkstra $et\ al.$ shortest path algorithm [65]. At this point, the active contours method is applied, incorporating image information, anatomical knowledge and geometrical constraints. From the image, the method retrieves histograms of each layer and computes a probability of given attenuation coefficient belonging to a layer. The anatomical information is included in the form of prior and it penalizes the crossing of boundaries and boundaries that correspond to abnormal layer thickness. The geometrical constraints penalize the contour length and avoid boundaries from folding back. All the boundaries are segmented simultaneously.

[2]The attenuation coefficients depend only on the tissue. Therefore, they are less influenced by some artifacts, such as the shadowing caused by blood vessels [64].

Rossant $et\ al.$ [66] proposed a novel active contours model based on the classical snakes model. The first step is the denoising of images through a nonlinear anisotropic diffusion filter. The initial boundaries are normally determined through a k-means segmentation and previous segmentations, since the boundaries are segmented in multiple stages. For the segmentation, the authors propose a parallel dual snake model. For individual boundary segmentation, this model is equivalent to the traditional snakes model, i.e., it uses edge information and smoothness constraints. Nonetheless, for simultaneous boundary segmentation of boundaries, the model adds an energy term that considers the parallelism between boundaries. The new energy term does not rely on anatomical knowledge of distances between boundaries (e.g., maximum, minimum and average distances), opposing to the other approaches mentioned so far. The parallelism is only controlled through a weight – a high weight encourages a fixed distance between boundaries, while a low value allows the distances between boundaries to vary along the curve. This is important because the thickness of retinal layers can change throughout the retina extent and as such, using distances based on simple anatomical knowledge may lead to incorrect segmentation of certain regions. To validate their method, the authors tested the method on images of healthy patients and patients with retinitis pigmentosa. In this disease, the outer segment (OS) thickness exhibits slight variations along the macula, usually linked to the degeneration of the photoreceptors [67].

3.1.3 Statistical modeling

The characteristics of each method of this section are highly dependent on the statistical model and how it is applied. Therefore, it is not possible to generalize any type of capability common to the methods based on statistical modeling.

Zhu $et\ al.$ [68] describe an approach based on Gaussian processes (GPs) designated as FloatingCanvas. Here, the heights of each surface are modeled by a GP. The covariance of the GPs is based on a Gaussian kernel, whose scale parameter controls the smoothness of the segmentation. The boundary segmentation is performed by the deformation of an initial surface according to a partial differential equation (similar to active contours methods). This equation is composed by a gravity term, an image term and a binary function. The gravity term pushes the surface to a specified direction, whereas the image term drives the surface to prominent features of the image (e.g., strong gradients or large intensities). The binary function selects which of these terms is activated. For instance, in the segmentation of the ILM, the image term is activated in regions where there are strong gradients, otherwise the gravity term is activated. The FloatingCanvas was only applied to ILM, ORNFL and IRPE, since these boundaries are relevant for the evaluation of glaucoma. The smoothness parameters are defined empirically, thus further development of this algorithm could benefit from learning these parameters from training data. Furthermore, these parameters could be learned for each spatial location of the retina, as opposed of being fixed for the whole retina extent.

Kajic $et\ al.$ [69] proposed a segmentation method based on active appearance model (AAM). Their algorithm initiates with a denoising method based on the dual-tree complex wavelet. Then, three boundaries with good contrast (ILM, IS-OS, BM) are sequentially segmented through adaptive thresholding. At this point, the AAM can

be constructed. The AAM is composed by shape and texture models. The former model is based on distances relative to the ILM. Whereas, the latter model is based on the mean and standard deviation of intensity of layers and the mean intensity of edge regions for multiple scales. The features also consider several positions along the horizontal direction, since the morphology of the retina varies throughout its extent. This increases the dimensionality of the feature vectors. To reduce the feature dimensionality, an inverse neural network is used. This method was selected because it can perform a nonlinear dimensionality reduction. The segmentation is performed by minimizing an objective function that compares the textures of the image with those predicted by the model. Moreover, the objective function also penalizes heavily segmentations of the ILM, IS-OS and BM that are considerably different from the preliminary results. The segmentation is finally refined by using AAM for each A-scan. One of the benefits of using a AAM is its capability of learning textural and shape variability of different regions of the retina, opposing to some other methods that use fixed smoothness constraints.

Rathke $et\ al.$ [70] proposed a statistical approach that exhibits some similarities to the method of Kajic $et\ al.$ [69]. Rathke $et\ al.$ propose a local appearance and shape models. The local appearance model represents the appearance of patches within layers and at the boundaries using a multidimensional Gaussian distribution. The shape model assumes that the positions of boundaries are also well represented by a multidimensional Gaussian distribution. For both models, the data is projected into a low-dimensional manifold by methods based on the principal component analysis (PCA). Furthermore, the covariance matrix estimation uses a regularization based on the $l1$-norm. The models are defined for each A-scan. The global models are assumed to be Markov random fields (MRF) and as such they can be obtained by the multiplication of the column-wise models. At the end, the global local appearance and shape models are combined and used in a variational inference approach to determine the segmentations. This work computes a global shape model that is more complete than the one proposed by Kajic $et\ al.$ [69][3]. The global shape is capable of learning morphological retinal alterations and typical amount of variability from training data, thus avoiding any type of regularizing post-processing.

3.1.4 Pixel classification

Pixel classification is another possible approach for the segmentation of retinal boundaries. In this approach, a classifier uses features to predict a class label for each pixel. Classifiers are simple to use and when properly applied can provide competitive results. Nevertheless, traditional classifiers lack regularization on its output and the features need to be designed for each application.

Szkulmowski $et\ al.$ [3] proposed a semi-automatic method for the segmentation of the hyperreflective complex composed by the OS and RPE. The features considered are mean and standard deviation of each pixel neighborhood. The method requires that the user selects a region of the OS+RPE complex (Fig. 3.6). The algorithm uses this selection to calculate the parameters of the feature model for the OS+RPE complex, which is assumed to be two-dimensional Gaussian function. Pixels are then classified into OS+RPE by selecting those that are within an elliptical region of the feature space that contains 99% of the pixels of OS+RPE, according to

[3]The model of Kajic $et\ al.$ [69] is computed for some sampled positions of the retina.

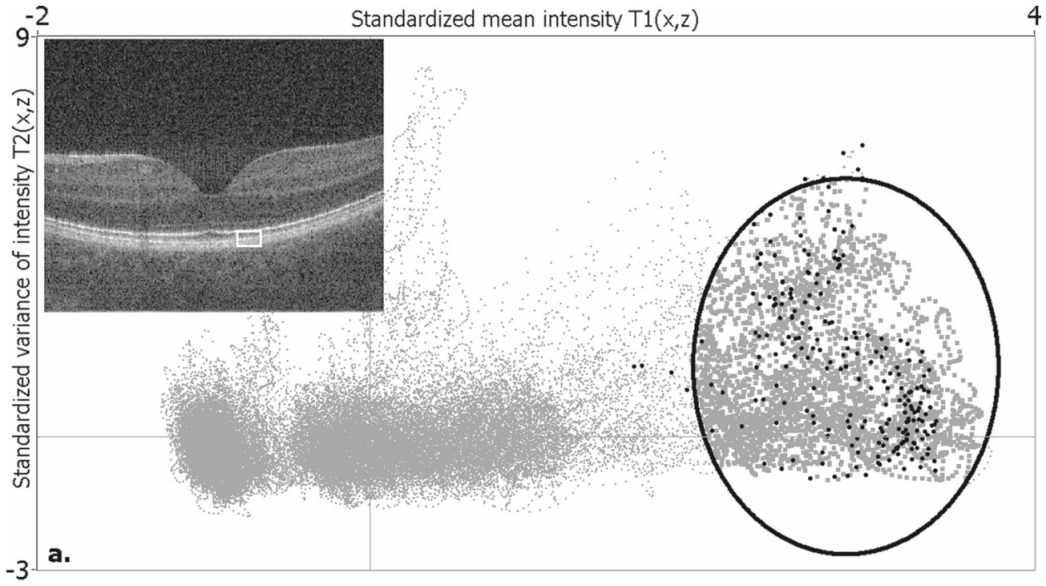


Figure 3.6: Details of the method proposed by Szkulmowski $et\ al.$ [3]. The white rectangle corresponds to the user defined selection. The black points are the pixels of the user defined selection in the feature space. The black ellipse contains 99% of the pixels of the OS+RPE according to the Gaussian model. The gray squares represent the pixels inside the ellipse, while the gray points correspond to the remaining pixels (image retrieved from [3]).

the Gaussian model (Fig. 3.6).

Vermeer $et\ al.$ [71] proposed an automatic method using the SVM classifier. The features consist of local averages and gradients of A-scans for 4 different scales. The output is then regularized by resegmenting the boundaries through an active contours method.

Kafieh $et\ al.$ [72] developed an unsupervised method that uses textural features based on the co-occurrence matrix, run-length matrix and other statistics. These features are nonlinearly transformed using diffusion maps. Afterwards, K-means is applied in the transformed space. The final segmentations are smoothed through interpolation.

The algorithm of Lang $et\ al.$ [62] pre-processes the images by applying an intensity normalization and a flattening. Thereupon, the algorithm uses a RF with spatial and local context aware features to compute the probability of a pixel belonging to a boundary. The spatial features consist in distances to fovea, to the inner boundary of the RNFL and to BM. The local context aware features are based on the intensity values of the neighborhood, derivatives of the outputs of a Gaussian filter bank and the mean vertical gradients of large neighborhoods below the considered pixel. The RF provided probability images that were used as input for the Canny edge detector [73] and a method based on the MSF [74]. The latter method attains better results because it incorporates regularizing constraints and segments boundaries simultaneously[4].

Antony $et\ al.$ [75] also use a RF classifier to provide probability images for the MSF method proposed in [76]. This method includes boundary and region terms, where a part of the classifiers detects boundaries, while

[4]Boundaries are segmented in 3 steps, each of those segments 3 boundaries simultaneously.

the other part detects layers. Most of the features are based on Haar, Gabor and Gaussian filter banks. The remaining features are statistical measures of the neighborhood, i.e., mean, variance, entropy, skewness and kurtosis.

3.1.5 Shortest path

Some methods use Dijkstra's shortest path algorithm [65] for the segmentation of retinal boundaries. This approach is simple and presents some robustness to the presence of noise and artifacts, since searching for the shortest path has a regularizing effect. On the other hand, this effect can cause smoothing of structures with a considerable curvature. This approach cannot segment boundaries simultaneously, thus each boundary is segmented in a particular search space, often defined by previous segmentations. Segmenting in this restricted search space avoids interference from the other boundaries.

One of the methods that uses the shortest path algorithm is the work proposed by Chiu $et\ al.$ [77]. This method starts by flattening the image, so the boundaries are approximately flat and the shortest path algorithm [65] can be applied. The flattening is based on a segmentation of the brightest pixels of each column in a smoothed image, which normally are located inside the RPE layer. Then, a graph is created for each image. The nodes have an one-to-one correspondence to the pixels of the image. The weights of the nodes are based on the gradient. In blood vessels regions, the weights are set to a constant to avoid coarse errors related to the low contrast of these regions. Thereupon, the first boundary is segmented by searching for the shortest path. The search space of each boundary is restricted by the previously segmented boundaries. At the end, some boundaries are resegmented in the fovea with a refined search space because some of them are very thin in this location.

Yang $et\ al.$ [78] proposed a method similar to the method of Chiu $et\ al.$ [77]. The method of Yang $et\ al.$ segments 9 boundaries – two more than the method of Chiu $et\ al.$ Additionally, the method of Yang $et\ al.$ defines the graph weights as a linear combination of features. Two of those features are the output of the Canny edge detector [73] and the vertical gradient. The information of gradient is computed in a coarse scale to improve the robustness of the method to artifacts and noise. Depending on the boundary, other features were used, such as edge direction and image intensity.

Later, the work of Chiu $et\ al.$ [77] was extended by Tian $et\ al.$ [79]. In this work, Tian $et\ al.$ use inter-frame information for the flattening and for the definition of search spaces. Using this information reduces computational complexity and improves the segmentation accuracy in some cases. The inter-frame information is only used for the boundaries that the authors deemed necessary. Another novelty is the use of biasing and masking, which are employed for segmenting boundaries that share the same search space. Biasing consists in multiplying the weight image (e.g., gradient) by a decreasing function, so the spatially lower boundary is segmented first (Fig. 3.7). Afterwards, the upper boundary is segmented using the masking operation to redefine the search space. Masking is just a further restriction of the search space by taking into account the segmentation of the previous boundary (Fig. 3.7). Overall, this work attains results that tend to be slightly worse than those of

Chiu $et\ al.$ [77]. Nonetheless, the algorithm of Tian $et\ al.$ [79] is about 8 times faster than one of Chiu $et\ al.$ [77][5].

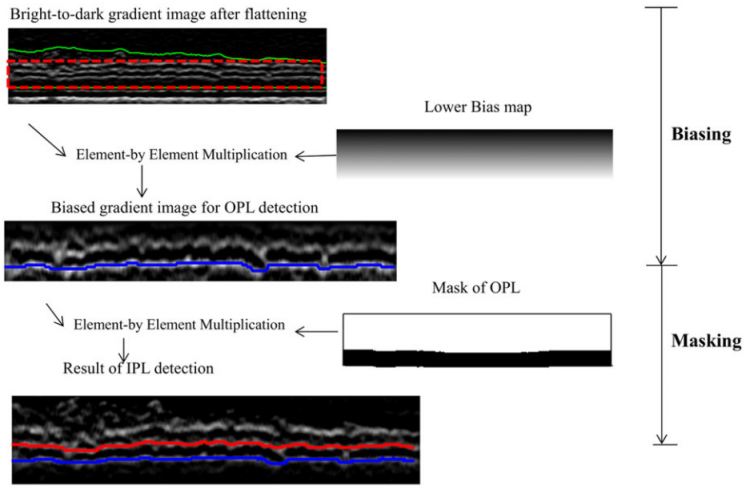


Figure 3.7: Example of biasing and masking operations. The red rectangle refers to the search space for segmenting the outer boundary of the outer plexiform layer (OPL) and the boundary between the inner plexiform layer and the inner nuclear layer (IPL-INL), shown as IPL in the image. The search space is biased and the outer boundary of the OPL is segmented (blue boundary). Afterwards, the search space is restricted using the masking operation and the IPL-INL is segmented (red boundary). Image retrieved from [79].

3.1.6 Multiple surface framework

Li $et\ al.$ [74] uses a graph-based multiple surface framework (MSF) to segment approximately parallel surfaces in layered tissues. This framework translates the multiple surface segmentation (MSS) problem into a minimum closed set problem, which can be efficiently solved through the minimum s-t cut/max-flow algorithm. Minimizing the energy terms of the MSS problem is equivalent to searching for the minimum cost surface, while being constrained by hard smoothness and interaction priors. The smoothness prior limits local variations of the surface, while the interaction prior restricts the distance between two boundaries.

The MSF was adapted to the segmentation of retinal boundaries by Garvin $et\ al.$ [76]. The segmentation method was modified to consider not only edge information, but also region information. Moreover, Garvin $et\ al.$ also added smoothness and interaction constraints that vary throughout the volume, since boundaries can present different behaviors depending on the retinal location (e.g., fovea). The limits of these constraints are learned from training data. The algorithm proposed by Garvin $et\ al.$ starts by flattening the OCT volume through a smoothed preliminary segmentation of the IRPE. Then, the retinal boundaries are segmented in two steps. The first step segments ILM and outer boundaries, since these present better contrast, while the second

[5]The data used for this comparison is based on different hardware, however this information provides some insight to the reduction of computational complexity.

step segments the inner boundaries.

Later, Song $et\ al.$ [80] extended the MSF by introducing the formulation for soft smoothness constraints. These constraints penalize local variations that diverge from the mean local variations learned from training data. In practical terms, these constraints control the rigidity of the surface. Increasing the rigidity reduces the influence of noise and artifacts, but also diminishes the importance of the image information. The segmentation scheme is the same as the one proposed by Garvin $et\ al.$ [76].

Dufour $et\ al.$ [1] introduced another type of constraint to the MSF: the soft interaction constraint. This type of constraint is similar to the soft smoothness constraint. The soft interaction constraints penalize distances that diverge from the mean distance between boundaries. In other words, this constraint tries to set the distances between surfaces to fixed values learned from training data. The soft interaction constraints improve the segmentation of surfaces that are spatially close and present similar gradient values. The structure of the algorithm is similar to what was proposed by Garvin $et\ al.$ [76]. The main differences are the segmentation of the ILM in a single stage and using that result as reference of the curvature of the retina, thus avoiding the use of flattening.

3.2 Quantification of drusen

The previous section presented several algorithms developed mainly for the segmentation of retinal layers in images of healthy patients[6]. This section will focus on algorithms developed for retinal volumes of patients with AMD and that perform the quantification of drusen.

The drusen material is limited by the ORPE and BM (Fig. 3.8 (a)). However, the ORPE is often not clearly defined because drusen material can be just as hyperreflective as the RPE (Fig. 3.8 (b)). Therefore, some algorithms presented in this section replace the quantification of drusen by the quantification of the space between IRPEDC[7] and BM in drusen regions (Fig. 3.8), since the ORPE is not visible in some cases. This is also a valid approach, since the total volume of the space between the IRPEDC and BM in drusen regions is also correlated with the risk of AMD progressing to an advanced stage [81].

Most of the algorithms that quantify drusen start by segmenting the ORPE (or the IRPEDC) and the BM. After the segmentation, these boundaries are used to detect drusen, typically by thresholding the distances between both boundaries (Fig. 3.9). Schlanitz $et\ al.$ [17] compared the segmentation of commercial systems with the manual detection of drusen by experts and concluded that the best algorithm only detects about 30% of the drusen with negligible error. The authors argue that interpolation applied to the RPE layer is probably the main cause for these results. This shows that even though commercial systems may produce accurate segmentations of retinal boundaries in volumes of healthy patients, they cannot handle well the presence of drusen. Therefore, algorithms robust to the presence of drusen or adapted for its detection are necessary to perform an accurate quantification of drusen.

[6] Some of the algorithms described were developed for diseases that alter just slightly the morphology of the retina, such as glaucoma.

[7] IRPEDC refers to the IRPE + drusen complex.

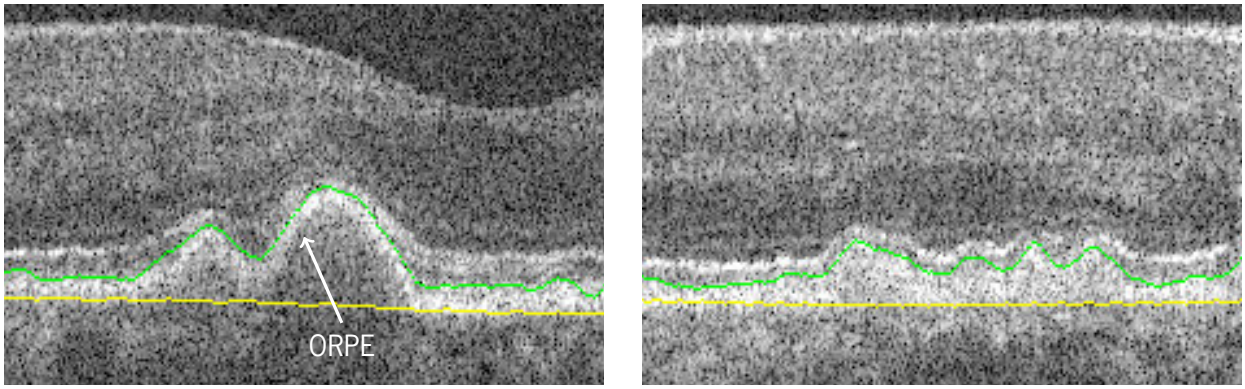


Figure 3.8: Examples of (a) drusen with a clearly defined ORPE boundary and of drusen where the ORPE is not visible. In the latter, the drusen material is just as hyperreflective as the RPE layer. The green boundary refers to the IRPEDC and the yellow boundary represents the BM.

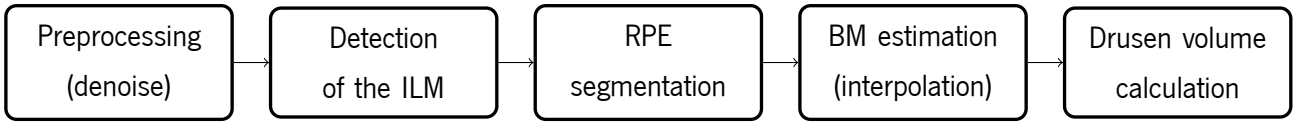
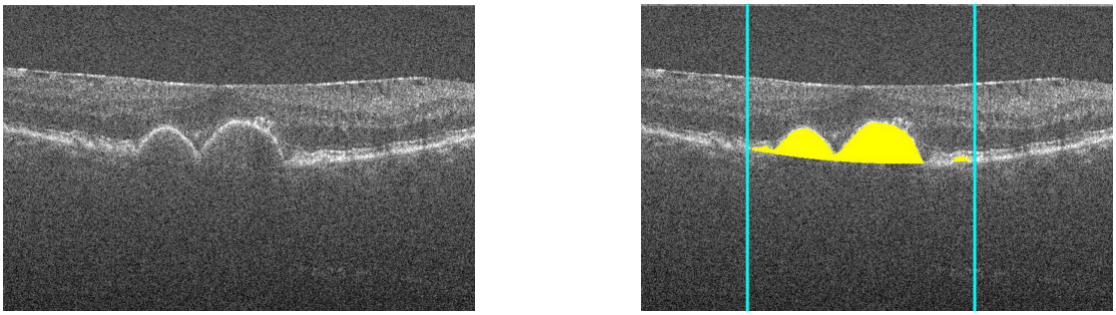


Figure 3.9: Diagram of the typical structure of a drusen segmentation algorithm. Farsiu $et\ al.$ [4] were possibly the first to propose this structure.

Recently, a few algorithms were specifically designed to automatically quantify drusen. One of the first was proposed by Farsiu $et\ al.$ [4] (Fig. 3.9). This algorithm starts by segmenting the ILM. This boundary is not essential for the quantification, but is important to spatially restrict the location of the RPE, which exhibits intensities similar to the RNFL. The ILM is detected by using a low pass filter (for noise attenuation), a vertical high pass filter (to detect the boundary) and some additional operations to remove outliers and ensure connectivity. Afterwards, in a noise filtered image, the space below this boundary is searched for the brightest locally connected pixels. This segmentation is used to obtain a rough contour of the ORPE boundary that is used as input to the active contours method. This method uses both information of the gradient vector flow and regularizing constraints to obtain a more accurate segmentation. To obtain the estimation of the BM, the ORPE segmentation is fitted to a 2^{nd} or 4^{th} order polynomial with a local convex condition. This is performed to smooth the ORPE elevations that possibly correspond to drusen. The space between the ORPE and BM corresponds to drusen material and so, the volume for the whole retina can be estimated by computing the drusen area of all slices. This work only presents visual results, but it seems to be able to provide accurate segmentations for large drusen (Fig. 3.10).

Other algorithms use a structure similar to the method proposed by Farsiu $et\ al.$ [4] (Fig. 3.9). One of which was proposed by Yi $et\ al.$[82]. This algorithm uses the ORPE segmentation obtained from Mujat $et\ al.$ [83]. The first step of the algorithm was the application of a Gaussian filter for noise reduction. Then, the ILM was detected by using an active contours algorithm based on gradient, where the initial input is based on a segmentation by



(a) **(b)**

Figure 3.10: Example of visual results of the algorithm of Farsiu $et\ al.$ [4]. In (a) a B-scan with drusen is presented and the respective segmentation (region of interest between the blue lines) is shown in (b) (images retrieved from [4]).

thresholding. Afterwards, the area above the RNFL is removed and it is applied an algorithm for joint anisotropic smoothing and edge-preservation. This algorithm was based on the minimization of energy functional, which corresponds to a problem of nonlinear optimization with no closed solution. So, the problem is solved by an iterative process that, in an alternating manner, optimizes **f** and **s**, a smooth and edge representations of the original image, respectively. In the energy functional, there are three main terms: smoothness constraint, data fidelity and edge control. The first allows the smoothing in **f** to occur only in zones where there are no edges and in a preferential direction (anisotropic nonlinear smoothing); the second penalizes larges differences between **f** and the original image; and the third controls the density of edges in image **s**. The ORPE boundary is obtained by thresholding the final images **f** and **s** and by retaining the lowest bright pixel of the two binarizations. The ORPE segmentation is then used to estimate the BM by fitting with a 2^{nd} order polynomial. Ultimately, the drusen volume is then calculated (just as in the algorithm of Farsiu $et\ al.$ [4]). Again, there are only visual results, nevertheless the algorithm seems to accurately segment drusen with considerable dimensions.

Gregori $et\ al.$ [39] proposed another algorithm that uses the same structure as Farsiu $et\ al.$ (Fig. 3.9) [4]. This proprietary algorithm of Carl Zeiss Meditec uses the built-in RPE segmentation method of the Cirrus HD-OCT (version of 2011) and a method for estimating the BM that is supposed to be included in the following generation of Cirrus HD-OCT. The details of this algorithm are unknown, as well as its accuracy results (only reproducibility results are available).

Another similar algorithm is proposed by Iwama $et\ al.$ [18] (Fig. 3.9). This algorithm was based on the built-in software of 3D OCT-1000 (by Topcon). Initially, a median filter is applied to attenuate noise. Then, a binarization is performed to obtain rough estimations of RNFL, OS and RPE. The following step uses the Canny edge detector [73] to determine the ILM and the IS-OS boundaries. Thereupon, the brightest line below the IS-OS boundary is considered as the IRPEDC boundary. This boundary is detected through global straight line fitting that starts in the lowest point of the rough RPE segmentation. Afterwards, the BM is determined by interpolation of the IRPEDC segmentation by fitting of 2^{nd} order polynomial. The final step is the drusen area calculation. The main goal of the article is to compare the drusen characteristics obtained by the algorithm in OCT and manually by expert graders in retinographies. When comparing drusen maximum size obtained by these modalities, the

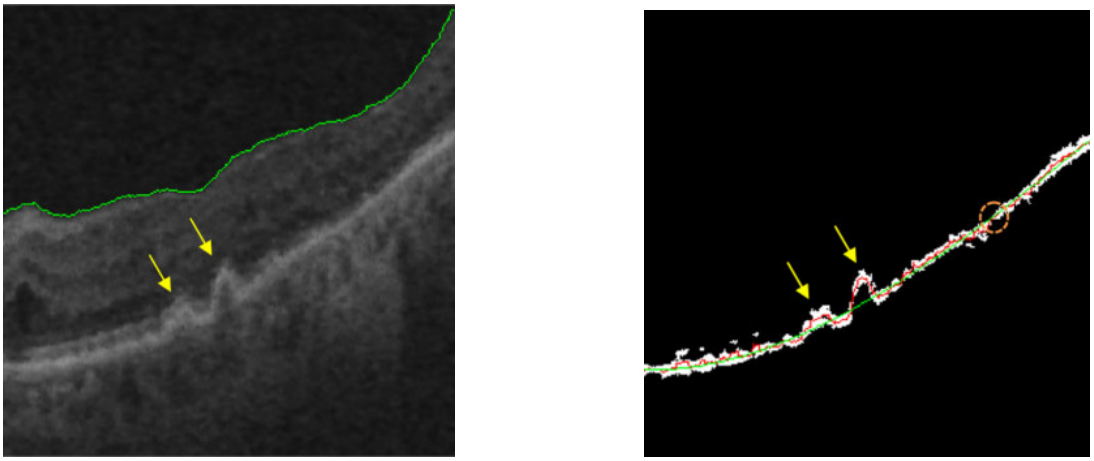
agreement is almost perfect, however in drusen area, only 8 out of the 18 eyes showed a good agreement (for a threshold for drusen height of 4 pixels). The graders reported algorithm failure[8] in about 69 images out of 2304. According to the authors, the algorithms fails mainly because it misses small drusen.

More recently, an algorithm following the same structure (Fig. 3.9) was developed by Chen $et\ al.$ [5]. Although, the structure is similar, there are a couple of significant differences in this algorithm. The first is the use of bilateral filtering (Fig. 3.11 (a)) which is a nonlinear filter that takes into account the geometric distance and intensity similarity of the image pixels to perform the smoothing. Then, the RNFL is extracted through thresholding. This is done in order to facilitate the segmentation of the RPE also by thresholding (Fig. 3.11 (b)). Afterwards, some operations are employed to remove small objects and to ensure the continuity of the RPE (Fig. 3.11 (b)). The BM is determined by fitting the middle axis of the RPE to a 3^{rd} order polynomial. The drusen area is considered to be between the middle axis of the RPE and the BM (Fig. 3.11 (b)). Another different aspect of this algorithm is the existence of a three step post-processing, to try to compensate its inaccurate RPE segmentation. Since the RPE segmentation is inaccurate, there will be elevations of the RPE middle axis (in relation to the BM) that will not correspond to drusen. The two first post-processing steps are aimed to remove these and other possible false positives. The first step is the removal of drusen present in a single slice, thus excluding drusen smaller than the azimuthal resolution. The second step tries to remove false positives through thresholding of shape and intensity features in an $en\ face$ projection. The third step consists in applying a 3D Gaussian filter to achieve smoother contour, which is consistent with drusen shape. The results of the algorithm were compared to an expert grader and reveal an absolute drusen area difference[9] of 19.97 $\pm$ 14.68 μm and an overlap ratio of 67 $\pm$ 9.14 %. The intergrader results have an absolute drusen area difference of nearly 9.98 $\pm$ 9.49 μm and an overlap ratio of 79 $\pm$ 9.65 %, which is better than the algorithm results. The absolute drusen area difference might be a misleading metric since it does not take into account the A-scans with drusen, where false positives might occur. Moreover, the author states that several drusen located on the image border are missed because of the of signal deterioration in these zones. Still, the algorithm might be useful in clinical practice and it evidences that the automatic drusen segmentation methods can achieve results comparable those of expert graders.

Baumann $et\ al.$ [8] proposed an algorithm with fewer steps based on a variation of OCT: the polarization-sensitive OCT (PS-OCT). The PS-OCT is sensitive to light polarization and it can measure the degree of polarization uniformity (DOPU). The RPE has a depolarizing character and so it presents a low DOPU. Thus, this layer can be easily segmented in a DOPU image by thresholding. This segmentation is not accurate and so, the authors suggest to pick for each column the pixel with lowest DOPU, which is certain to belong to the RPE. Afterwards, the RPE middle axis is obtained through the use of some interpolation operations, which remove outliers and missing points. To calculate the BM, several consecutive Savitzky-Golay filters (3^{rd} order polynomial) are applied to the RPE middle axis. The effect of these filters is equiparable to an interpolation operation [84]. The algorithm is reported to detect 96.5 % of the drusen without significant error and its segmentation area has a difference of

[8]The algorithm failure is reported as obvious disruption of the segmentation for more than 20% of the columns of each B-scan or just 5% if the disruption occurs consecutively [18].

[9]Absolute drusen area difference corresponds to the amount of error per A-scan in the A-scans that have drusen.



(a) (b)

Figure 3.11: Example segmentation of the RPE with the algorithm of Chen $et\ al.$ [5]. The original image is smoothed by using a bilateral filtering (a) and then the RPE is segmented by thresholding (b). Afterwards, the middle RPE axis (red line in (b)) is calculated and the BM (green line in (b)) is estimated. The yellow arrows point to drusen and the orange circle is a location that was not connected after the segmentation of the RPE and was reconnected after some operations (images retrieved from [5]).

6% compared to the the manually segmented area. The main causes of failure of the algorithm are the presence of depolarizing material in the drusen, lack of RPE (i.e., GA) and lateral signal deterioration. Even though the PS-OCT enhances the contrast of RPE, facilitating drusen segmentation, this technology is not widely available. Thus, this algorithm cannot be applied as a general clinical practice (at least not yet).

Chiu $et\ al.$ [85] proposed a distinct approach to segment drusen. This approach is based on the graph search method proposed by the same first author [77], described previously in section 3.1.5. The authors mention that several changes were made to adapt the algorithm to AMD characteristics, namely the presence of drusen and GA. One of the main changes was the redefinition of the search spaces of the IRPEDC and the BM using preliminary segmentations. This strategy allows the segmentation of boundaries without interference of nearby transitions. The first step of the algorithm is image downsampling to reduce the computation time. In this algorithm, only three boundaries are detected: ILM, IRPEDC and BM. The segmentation of boundaries is also based on Dijkstra's shortest path algorithm [65] and the graph's weights are also based on the vertical gradient of the image, except for the BM that needs the incorporation of geometrical distances for its regularization. A preliminary segmentation of the ILM, IRPEDC and BM boundaries is performed by thresholding an image smoothed with a Gaussian filter. The convex hull of the preliminary segmentation of the IRPEDC is then used to flatten the image. After the flattening, the ILM is resegmented using information of the preliminary segmentations to restrict the search space. Thereupon, the IRPEDC is resegmented also using the information of the preliminary segmentations. This segmentation is used as the upper limit of the search space defined for the subsequent segmentation of the BM. Afterwards, the pixels with highest intensity of the RPEDC are segmented using the BM as the lower limit of the search space. This segmentation is used as the lower limit of the final IRPEDC

segmentation, while the upper limit is a 10 μm[10] offset from the lower limit. This algorithm was not used to detect drusen, it only segmented their limiting boundaries, thus results of drusen detection are not reported. Nonetheless, the authors evaluated the mean thickness error ($\pm$ standard deviation) of the RPEDC and the percentage of A-scans with errors higher than 5 pixels. For the first metric, the intergrader variability was 4.5 $\pm$ 3.5 μm and for the algorithm was 3.2 $\pm$ 2.6 μm. As for the second metric, the intergrader variability was 11.3 % and for the algorithm was 7.5%. The algorithm presents better results than the intergrader variability. The results are not discriminated for drusen regions, so it not possible to evaluate the accuracy of the segmentations in these locations.

The method proposed by Dufour $et\ al.$ [1] that was described in section 3.1.6 was also adapted to volumes of patients with AMD. The framework is exactly the same, but the structure of the algorithm is altered after the segmentation of the ILM. For healthy retinas, the scheme consists in a simultaneous segmentation of the IS-OS and BM boundaries. For retinas with drusen, they divide the algorithm into a few steps: coarse segmentation of the BM and the IS-OS, individual segmentation of the BM and simultaneous segmentation of the IS-OS and the IRPEDC (the latter is used as an auxiliary boundary). The authors justify this alteration mainly because the BM segmentation would follow the ORPE in the cases of large drusen. This is partially explained by the lack of contrast of the BM in those regions (Fig. 3.1), as referred by Dufour $et\ al.$ The ORPE is segmented instead of the BM because the MSF still retains parallelism and a near constant distance relative to IRPEDC in drusen regions. The results in images with drusen were reported in the form of mean unsigned error ($\pm$ standard deviation) for each surface. The algorithm results were 3.25 $\pm$ 2.82 μm and 5.03 $\pm$ 5.74 μm, while the intergrader results were 3.51 $\pm$ 2.37 μm and 5.39 $\pm$ 4.64 μm for the IS-OS boundary and the BM, respectively. The algorithm results are better than the intergrader results, nonetheless the metric used does not state the amount of error for drusen regions and the IS-OS boundary is detected instead of the IRPEDC[11]. The authors also present a plot of the error of IS-OS and BM segmentations in drusen locations, which evidences an increase of errors in large drusen. The errors for the IS-OS exceed slightly the 10 μm in large drusen. As for the BM, the error goes from less than 1 μm for very small drusen to about 80 μm for large drusen (over 200 μm). This error rate in large drusen for the BM is still justified by the lack of contrast of the BM in drusen and the parallelism imposed by the MSF.

The previous method did not detected drusen location, however the same main author proposed another algorithm for that purpose (Fig. 3.12) [6]. The BM segmentation of the previous algorithm [1] is used as reference for selecting sampling positions, where features are computed. The features are based on the Leung and Malik filter bank[12], which is commonly used for texture analysis [86]. After calculating the features, they are inserted in a RF classifier that classifies each column as drusen or non-drusen. The area under the receiver operating characteristic (ROC) curve of this algorithm is 0.94. Drusen locations can be combined with segmentations of

[10] 10 μm corresponds approximately to about 3 pixels.

[11] If these segmentations were used for calculating areas, these would include a part of the PL.

[12] The Leung and Malik filter bank is composed by Gaussian filters, 1^{st} and 2^{nd} derivatives and Laplacian of the Gaussian function [86].

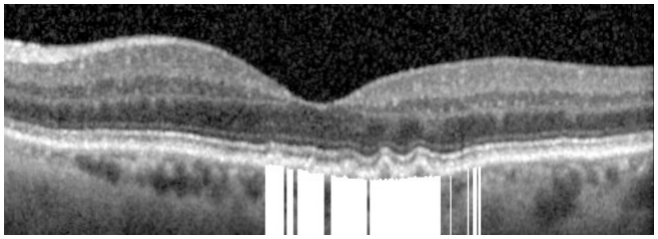


Figure 3.12: Example of drusen location detection by using the algorithm of Dufour $et\ al.$ [6]. The white lines at the bottom of the image signalize the presence of drusen in that column (image retrieved from [6]).

drusen's limiting boundaries to compute the total area or volume of drusen material. The locations of drusen can also be employed for a differentiated segmentation of these regions.

More recently, Novosel $et\ al.$ [87] adapted an active contours method to images of patients with AMD. This method is based on the active contours method for healthy patients proposed by the same main author [63], which was already described in section 3.1.2. The authors performed two major modifications to adapt the method to images of patients with AMD. The first was adding BM to the segmentation. In healthy regions, the ORPE and the BM are indistinguishable, however in drusen regions they can be differentiated (Fig. 3.1). The BM is the only boundary that is segmented solely with the procedure to determine the initial boundaries, which is based on Dijkstra $et\ al.$ shortest path algorithm [65]. The preliminary segmentations of other boundaries are used as the initial contours of the active contours method. The other major modification was adding intensity information and prior knowledge to the graph weights used when determining the initial segmentations. The intensity information creates a slight bias towards the interface of layers, since the intensity helps differentiate the transitions within layers from the actual interfaces of layers. The prior knowledge is introduced by varying the weights that control the cost of horizontal and diagonal edges. For instance, the BM is a very flat boundary, so this information can be introduced by setting a high horizontal weight, which lowers the cost of horizontal edges. In terms of results, the mean unsigned errors of the ORPE and BM were 10.5 and 5.5 μm, respectively. As for the overlap of drusen areas, the dice coefficient was 0.81[13].

3.3 Summary

This chapter reviewed methods used for segmenting retinal boundaries in OCT images of healthy patients and of patients with AMD. For images of healthy patients, there is a wide variety of segmentation methods. These methods can range from peak search methods to more elaborate methods, such as the MSF or active contours methods. The most elaborate algorithms use prior knowledge and smoothness constraints to improve their robustness to noise and artifacts. These sources of information are easier to incorporate when segmenting images of healthy patients than in images of patients with diseases (e.g., AMD), which considerably alter the retinal morphology. For healthy patients, the boundaries are smooth and there are no morphological changes

[13]These results are relative to 30 B-scans of 15 different patients with AMD.

caused by diseases. This explains the considerably lower number methods proposed for retinas of patients with AMD. The main morphological changes related with intermediate AMD are the appearance of drusen and GA. The quantification of drusen can be used to assess the risk of AMD progressing to an advanced stage. Thus, automatic methods for quantifying drusen are necessary. The quantification of drusen can be performed by segmenting their limiting boundaries and determining the locations of these lesions. For the segmentation, the methods need to take into consideration that the ORPE and BM are distinguishable in drusen regions, as opposed to what occurs in healthy regions. Moreover, the methods have to carefully apply the smoothness constraints, since these can oversmooth drusen, which correspond to elevations of the RPE layer. For detecting drusen, most approaches simply apply a threshold to the distances of the limiting boundaries or detect all the drusen material of the retina without detecting individual drusen. Dufour $et\ al.$ [6] proposed an alternate approach, where drusen are detected through a classification based on textural features. In general, the authors report that the major problems of quantifying drusen are: missing small drusen, oversmoothing of limiting boundaries and the attraction of the BM segmentation to the ORPE. In the next chapters, it is proposed a method that tries to address these issues.

Chapter 4

Theoretical background

This chapter will expose the theoretical background required to understand both the proposed segmentation method for boundaries limiting drusen and the proposed method for detecting drusen.

The proposed segmentation method is based on the multiple surface segmentation (MSS) problem, which is an optimization problem with the aim of determining the surfaces with the minimum cost and, at the same time, feasible according to regularizing and spatial constraints [76].

The proposed detection method extends a drusen detection method proposed by Dufour $et\ al.$ [6]. This method detects drusen by using a RF to classify each column as drusen or non-drusen. In this work, it is proposed an extension of this method that incorporates local information into the classification. This information is added by combining the RF with a HMM, thus creating a hybrid HMM-RF model. The theoretical aspects of this hybrid model will be exposed in section 4.2

4.1 Extended Multiple Surface Segmentation

The MSS problem can be solved using the MSF, which was progressively developed by several authors (as mentioned in section 3.1.6). For that, the MSF formulates other type of problems that are equivalent to the MSS problem. Therefore, this section will start by outlining the problems supporting the MSF (subsection 4.1.1) and, only then, it will detail the MSS problem and how the MSF is used to solve it (subsection 4.1.2). In this work, it is proposed an extension of the MSS problem that includes local shape priors. The extension of this problem and the subsequent extension of the MSF will be presented in subsection 4.1.3.

4.1.1 Base problems for the MSF

As mentioned earlier, the MSS problem has the goal of finding the surfaces with the minimum cost under regularizing and spatial constraints [76]. This problem can be solved using the MSF, which formulates a minimum closed set problem that is equivalent to the MSS problem. The MSF can also formulate a minimum s-excess problem instead, if the MSF formulation requires the existence of outgoing arcs from the minimum closed set

(Fig. 4.1). This means that when using soft smoothness or interaction constraints, the MSF formulates a minimum s-excess problem, since these constraints require arcs leaving the minimum closed set (more details in subsections 4.1.2.2 and 4.1.2.3). Both of these problems can then be converted to the minimum s-t cut problem (Fig. 4.1) [74]. Such conversion is useful because there are very efficient algorithms for solving the minimum s-t cut problem in the literature, such as the Boykov-Kolmogorov (B-K) [88]. The next subsections will describe the minimum closed set and the minimum s-excess problems, as well as the respective conversions of these problems to the minimum s-t cut problem.

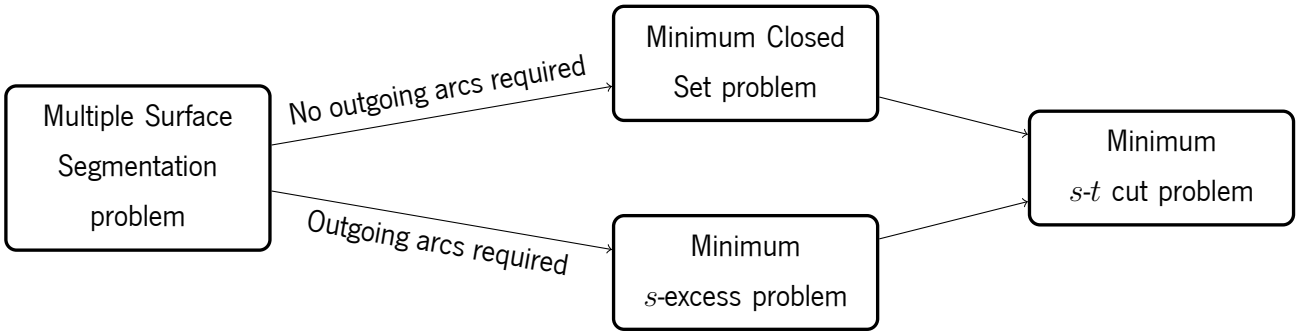


Figure 4.1: Diagram of the conversions of the minimum surface segmentation problem so it can be solved efficiently.

4.1.1.1 Minimum closed set problem

To explain the minimum closed set problem, first it is necessary to define a graph $G = (N, A)$, where N is the set of nodes and A the set of arcs composing the graph. A closed set (or closure) is a subset of nodes $N_c \subseteq N$ that has no outgoing arcs (Fig. 4.2).

The minimum closed set problem consists in finding the closed set whose node weight sum is minimal [89]- This can be formulated as [90]:

$$\min \quad \sum_{i \in N} w_i x_i, \tag{4.1}$$

$$s.t. \quad x_i \leq x_j, \quad \forall (i,j) \in A, \tag{4.2}$$

$$x_i \in \{0,1\}, \quad \forall i \in N, \tag{4.3}$$

where w_i and x_i are respectively the weight and label associated with each node i. When $x_i = 1$, the node i is in the minimum closed set, on the other hand, if $x_i = 0$ then node i does not belong to the minimum closed set. The Eq. 4.2 encodes the information that if a node i belongs to the closure, then its successors must also be in the closure. In other words, there are no outgoing arcs in the minimum closed set.

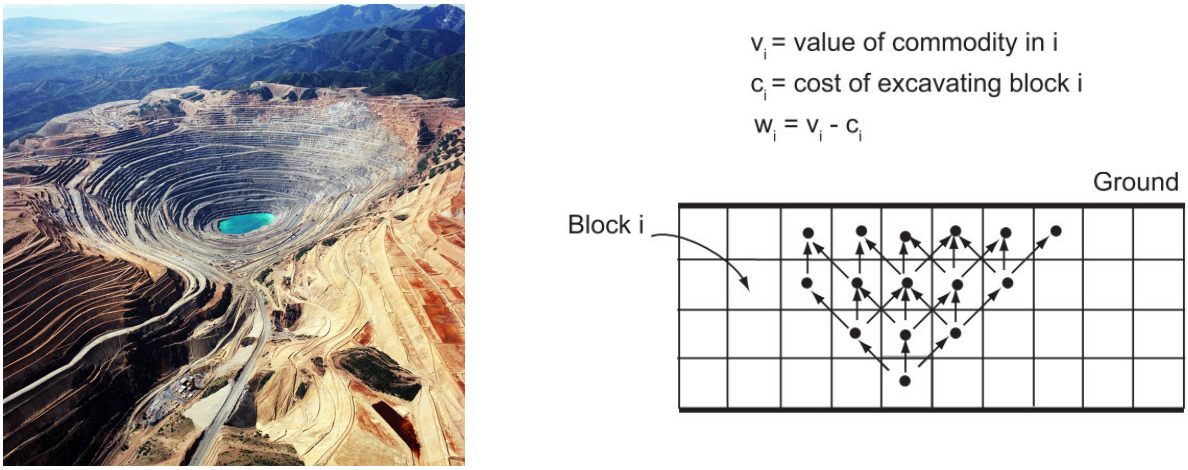
One practical application of the minimum/maximum closed set problem[1] is designing the geometry of an open pit mine (Fig. 4.3 (a)). In an open pit mine, a piece of land is progressively excavated to retrieve ore, which contains minerals of interest. The land is divided into blocks and each block has a value that depends on

[1]The maximum closed set problem can be converted into a minimum closed set problem by negating the node weights. [91].

(a) Closed set **(b)** Not a closed set

Figure 4.2: Examples of a (a) closed set and a (b) set that is not closed (image retrieved from [91]).

the amount of minerals of interest in them[2] (Fig. 4.3 (b)). Nevertheless, to retrieve the ore of each block there is a cost of excavation. Furthermore, several blocks are only accessible after other blocks are excavated (e.g., one block beneath another block). This means that some blocks might not contain minerals that justify their excavation, but they might still be excavated to access other blocks. Therefore, the geometry of the mine must be defined to maximize the profit of retrieving ore. This problem can be formulated as the maximum closed set by defining nodes weights w_i as the profit of each block and by using directed arcs to represent the relations of precedence for excavating the blocks, i.e., a directed arc from i to j establishes that for excavating block i, it is necessary to excavate block j first (Fig. 4.3). By formulating the problem in this manner, the geometry of the mine must be a closed set, otherwise there would be outgoing arcs, which would represent a physically impossible situation where some blocks are excavated without being accessible.



(a) Open pit mine photograph **(b)** Open pit mine schematic

Figure 4.3: Example of an open pit mine. The image (b) contains a graph that exemplifies how the problem of defining the geometry of an open pit mine can be formulated as maximum/minimum closed set problem. Image (a) was retrieved from [92] and (b) was adapted from [91].

[2]The value of each block is estimated using geological information obtained from drill cores [91].

4.1.1.1.1 Conversion to the minimum s-t cut problem

As mentioned previously, the minimum closed set problem can be transformed into a minimum s-t cut problem [89]. This transformation is performed because there are very efficient algorithms to solve the minimum s-t cut problem, such as the Boykov-Kolmogorov (B-K) algorithm for the scope of image processing[3] [88].

The minimum s-t cut problem is graphically represented by graph $H_{st} = (V_{st}, E_{st})$ (Fig. 4.4). The node set V_{st} is composed by the source (s) and sink (t) nodes along with the nodes in set V, which is related to V_{st} as follows: $V_{st} = V \cup \{s\} \cup \{t\}$ [91]. The arc set E_{st} has 2 types of arcs: t-links and n-links [88]. The t-links are those that link the nodes of V to the nodes s and t (solid lines in Fig. 4.4) [88]. The n-links refer to the inner connections inside V (dashed lines in Fig. 4.4) [88]. A s-t cut is a set of arcs removed from E_{st} that partitions the node set V into the sets S and T. The sets S and T are composed by the nodes of V that are connected directly or indirectly to nodes s and t after a s-t cut, respectively. The partition of V must also obey the following rules: $S \cup T = V$ and $S \cap T = \emptyset$.

The minimum s-t cut problem can be used directly for image segmentation (Fig. 4.4). In that context, the nodes of V represent the pixels and the capacities of the s-links and the t-links are related with the probability

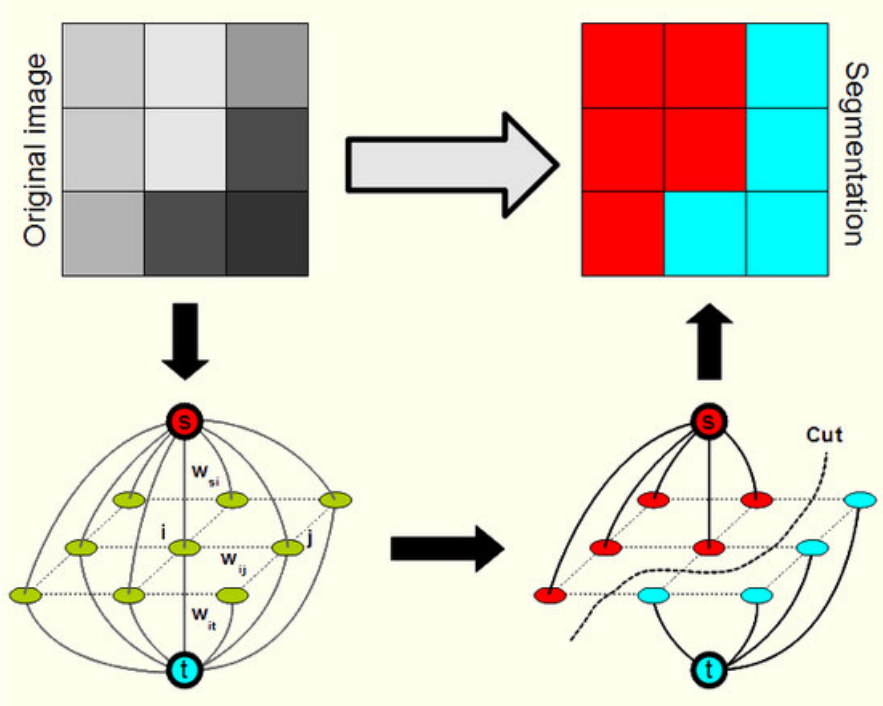
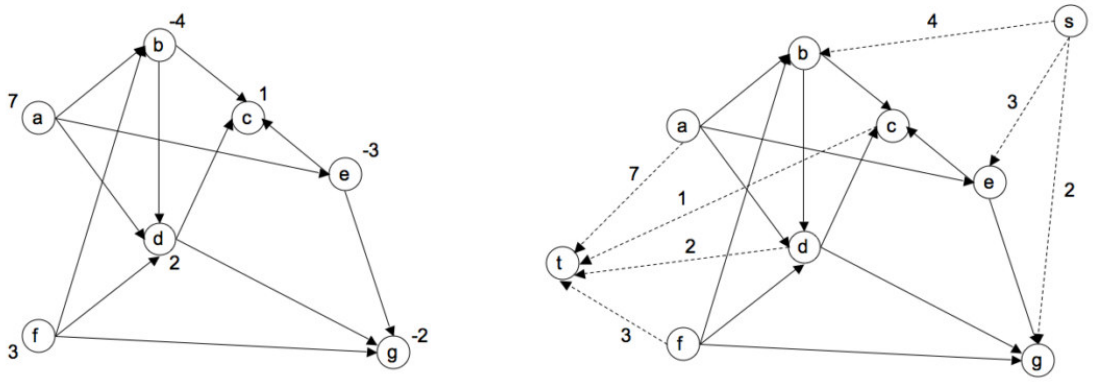


Figure 4.4: Example of how the minimum s-t cut problem can be used for image segmentation. In the left graph, the nodes of N are depicted in green and the n-links are represented by dashed lines. The t-links are presented as solid lines between the nodes of N and terminal nodes (s and t). The sets S and T are depicted in red and cyan, respectively (image retrieved from [93]).

[3]The image processing problems usually result in grid graph. For this type of graphs, the B-K algorithm is generally an efficient algorithm [88].

of a pixel belonging to different classes (e.g., foreground or background) [88]. Additionally, the capacities of the n-links can be interpreted as a cost of having a discontinuity between two pixels [88]. Normally, this cost is related with the similarity between pixels [88], i.e., the cost of a discontinuity is large for pixels with similar properties (e.g. intensity) and low for those with very different properties (Fig. 4.4).

To transform the minimum closed set problem into the minimum s-t cut problem, it is necessary to convert graph $G = (N, A)$ into a s-t graph $G_{st} = (N_{st}, A_{st})$ (example in Fig. 4.5). The new node set N_{st} is equal to N, but it also includes the s and t nodes. The set A_{st} incorporates the same arcs as A (solid arcs in Fig. 4.5). In the minimum closed set problem, the arcs A have no capacities, however for the A_{st} these arcs have the capacities set to ∞. Furthermore, the A_{st} includes new arcs from the node s to each node $i \in N^-$ and from each node $j \in N^+$ to the node t (dashed arcs in Fig. 4.5). The set N^- is composed by the nodes of set N with negative node weights ($w_i < 0$), while the set N^+ is composed by nodes of set N with positive node weights ($w_j > 0$). The capacities of the arcs originating in s are set as $-w_i$, while the capacities of the arcs terminating in t are defined as w_j (example in Fig. 4.5).



(a) Minimum closed set problem – graph G

(b) Minimum s-t cut problem – graph G_{st}

Figure 4.5: Conversion of the graph of the minimum closed set problem (G) into the graph of the minimum s-t cut problem (G_{st}). The arcs without a number associated have a capacity of ∞ (image adapted from [91]).

.

For a finite s-t cut of graph G_{st}, the node set S is always a closed set [91]. If this was not true, it would imply that there is at least an arc from set S to set T, i.e., an arc (i, j), given that $i \in S$ and $j \in T$. In graph G_{st}, only the n-links could comply with these requirements, however these cannot be cut because they have an infinite capacity. Therefore, the a s-t must only include t-links. More concretely a s-t cut of graph G_{st} is composed by arcs from s to nodes in T and arcs from nodes of S to t [91, 90]:

$$C(s \cup S, t \cup T) = \sum_{(s,\, j) \in A_{st},\ j \in T} c_{sj} + \sum_{(i,\, t) \in A_{st},\ i \in S} c_{it}, \tag{4.4}$$

where c_{sj} is the capacity of an arc from the source node s to a node j and c_{it} is the capacity of an arc from a node i to the sink node t. The Eq. 4.4 can be converted for the scope of the minimum closed set problem using

the respective node weights [91, 90]:

$$C(s \cup S, t \cup T) = \sum_{i \in T \cap N^-} -w_i + \sum_{i \in S \cap N^+} w_i \tag{4.5}$$

$$= \left(\sum_{i \in N^-} -w_i - \sum_{i \in S \cap N^-} -w_i \right) + \sum_{i \in S \cap N^+} w_i \tag{4.6}$$

$$= \sum_{i \in N^-} -w_i + \left(\sum_{i \in S \cap N^-} w_i + \sum_{i \in S \cap N^+} w_i \right) \tag{4.7}$$

$$= \sum_{i \in N^-} -w_i + \sum_{i \in S} w_i. \tag{4.8}$$

The term $\sum_{i \in N^-} -w_i$ is just a constant that represents the sum of all weights of N^- and can be disregarded for optimization purposes. With this, it possible to conclude that finding the minimum cost s-t cut is equivalent to minimizing the sum of weights of S. Since S is a closed set, then this set corresponds to the minimum closed set (Fig. 4.6).

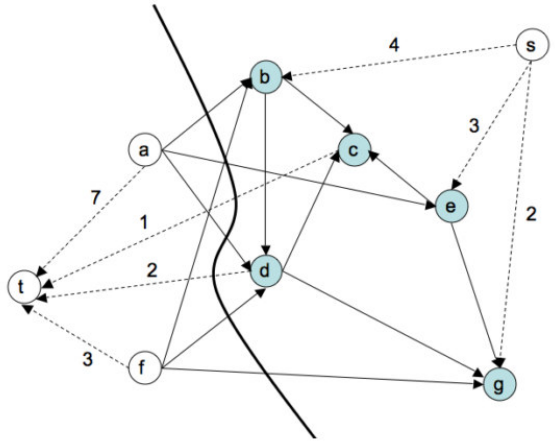


Figure 4.6: Relation between the minimum s-t cut problem and the minimum closed set problem. The thick line represents the minimum s-t of the graph and the blue nodes represent the set S. This set corresponds to the minimum closed set of the original graph from the minimum closed problem (Fig. 4.5 (a)). The arcs without a depicted capacity have a capacity of ∞ (image adapted from [91]).

4.1.1.2 Minimum s-excess problem

The minimum s-excess problem is a relaxation of the minimum closed set problem [94], which was presented in subsection 4.1.1.1. The problem considers a graph $G^s = (N^s, A^s)$, where N^s is the node set and A^s is

the arc set. The problem is formulated as [91, 94]:

$$min \quad \sum_{i \in N^s} w_i x_i - \sum_{(i,j) \in A^s} u_{ij} v_{ij}, \tag{4.9}$$

$$s.t. \quad x_i - x_j \leq v_{ij}, \qquad \forall (i,j) \in A^s, \tag{4.10}$$

$$x_i \in \{0,1\}, \quad \forall i \in N^s, \tag{4.11}$$

$$v_{ij} \in \{0,1\}, \quad \forall (i,j) \in A^s, \tag{4.12}$$

where w_i is the weight associated with node i and x_i is the binary variable that indicates if a node is in the minimum s-excess set. The variable v_{ij} is a binary variable that indicates if a penalty equal to the capacity of the arc u_{ij} is paid so node i is in the minimum s-excess set and its successor j is not. In practice, if $v_{ij} = 1$ the condition of closed set is relaxed and the minimum s-excess set may include node j. This contrasts with the minimum closed set problem, in which the successors of any node in closed set must also belong to closed set (section 4.1.1.1).

4.1.1.2.1 Conversion to the minimum s-t cut problem

The maximum s-excess can also be converted into a minimum s-t cut problem [94]. For the conversion, the graph of the minimum closed set problem G^s must be converted into graph G^s_{st}. This graph is similar to graph G_{st} used for the conversion of the minimum closed set problem (section 4.1.1.1.1). The only difference is that the capacities of the arcs in A^s are not reset to ∞ for the arc set A^s_{st}.

A s-t cut of G^s_{st} is given by [91, 94]:

$$C(s \cup S, t \cup T) = \sum_{(s,\, j) \in A^s_{st},\, j \in T} c_{sj} + \sum_{(i,\, t) \in A^s_{st},\, i \in S} c_{it} + \sum_{i \in S,\, j \in T} u_{ij}, \tag{4.13}$$

where c_{sj} is the capacity of an arc from the source node s to a node i, c_{it} is the capacity of an arc from a node j to the sink node t and u_{ij} is the capacity of an arc (i,j), given that node i belongs to S and node j to T.

A s-t cut of G_{st} can also be defined using the node weight and arc capacities of the minimum s-excess problem [91, 94]:

$$C(s \cup S, t \cup T) = \sum_{i \in T \cap N^-} -w_i + \sum_{i \in S \cap N^+} w_i + \sum_{i \in S,\, j \in T} u_{ij} \tag{4.14}$$

$$= \left(\sum_{i \in N^-} -w_i - \sum_{i \in S \cap N^-} -w_i \right) + \sum_{i \in S \cap N^+} w_i + \sum_{i \in S,\, j \in T} u_{ij} \tag{4.15}$$

$$= \sum_{i \in N^-} -w_i + \left(\sum_{i \in S \cap N^-} w_i + \sum_{i \in S \cap N^+} w_i \right) + \sum_{i \in S,\, j \in T} u_{ij} \tag{4.16}$$

$$= \sum_{i \in N^-} -w_i + \sum_{i \in S} w_i + \sum_{i \in S,\, j \in T} u_{ij}, \tag{4.17}$$

where the term $\sum_{i \in N^-} -w_i$ is a constant, which does not interfere with the optimization. Therefore, finding the minimum s-t cut in G^s_{st} is equivalent to searching for the minimum s-excess set in G^s.

4.1.2 Multiple surface segmentation problem

As referred before, the objective of the MSS problem is to find the surface with the minimum cost under regularizing and spatial constraints. In this work, the MSS problem is extended to also include local shape priors, but those will only be addressed in section 4.1.3. The MSS problem can be formally represented by the minimization of the energy function of Eq. 4.18. The external boundary energy term (E_{bound}) incorporates information from the image into the segmentation. Moreover, the smoothness energy (E_{smooth}) is responsible for the regularization of the surfaces and the interaction energy (E_{inter}) controls the distances between surfaces segmented simultaneously.

$$E(\boldsymbol{S}) = \sum_{i=1}^{n}(E_{bound}(S_i) + E_{smooth}(S_i)) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E_{inter}(S_i, S_j), \tag{4.18}$$

where the set $\boldsymbol{S}$ is composed by n surfaces. The n value refers to the number of surfaces segmented simultaneously with the MSF. S_i is the surface that corresponds to boundary i (e.g., IRPEDC or BM). The MSF translates the MSS problem into a minimum closed set problem (or minimum s-excess problem) by creating a graph $G = (N, A)$ specific for that purpose. For each surface S_i, the MSF defines a node subset N_i ($N_i \subseteq N$) as search space (Fig. 4.7 (b)). A single node of the subset N_i is represented as $N_i(x, y, z)$, having a one-to-one correspondence to the voxels of the volumetric image $I(x, y, z)$ (Fig. 4.7 (b)). By considering the image I to be composed by columns perpendicular to the xy plane, a surface can be defined as a height function $f_i(x, y)$ that passes exactly once through each column at $(x, y, f_i(x, y))$, as exemplified in Fig. 4.7 (a). The next subsections will further explain each energy term of the energy function of the MSS problem and describe how the MSF translates them into the graphical representations.

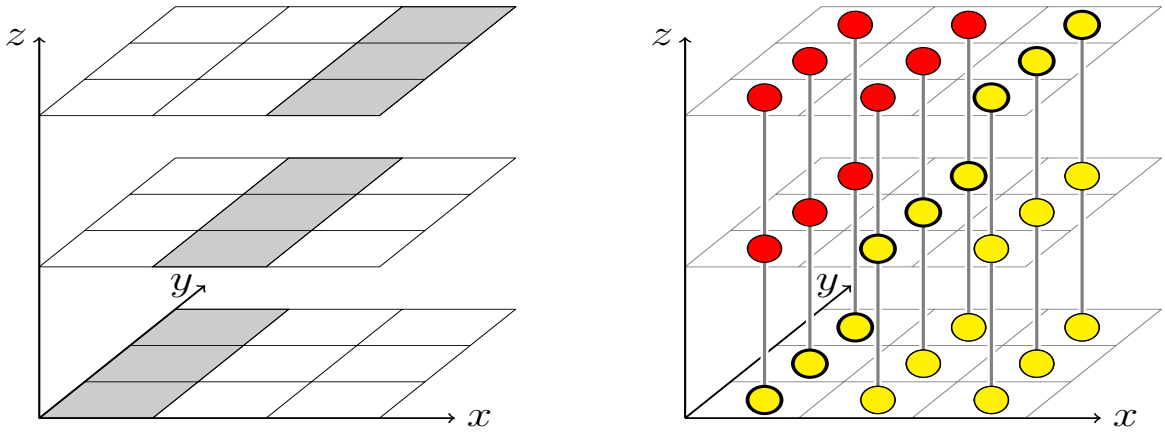
4.1.2.1 External boundary energy

The external boundary energy term of a given surface S_i is defined as:

$$E_{bound}(S_i) = \sum_{x\in X}\sum_{y\in Y}\gamma_i(x, y, f_i(x, y)), \tag{4.19}$$

where $f_i(x, y)$ is the height function of surface S_i and the sets X and Y include all the possible values of x and y in node set N_i. For the context of OCT images, X will include all the columns of each B-scan and the set Y will contain all the B-scans. Moreover, the weights $\gamma_i(x, y, z)$ are inversely proportional to the probability of node $N_i(x, y, z)$ belonging to surface S_i. These weights are frequently based on characteristics extracted from the image (e.g., gradient).

Without the smoothness and interaction constraints, the MSS problem is simplified to searching for the surface with the minimum cost. This corresponds to finding the surface with minimum sum of weights $\gamma_i(x, y, z)$ (Fig. 4.8 (a)), which is the same as minimizing E_{bound}. This minimization pushes each surface S_i to prominent features of the image. For instance, when E_{bound} is based on the gradient of retinal OCT images, then the surface S_i is pushed towards the boundaries between retinal layers.



(a) Surface S_i as height function $f(x, y)$ (depicted in gray) **(b)** Surface S_i as a node set N_i with different labels

Figure 4.7: Representation of surface S_i as a height function (a) and as a node set (b). The grid structure corresponds to the volumetric image I and the node set of (b) to N_i. The (x, y) columns are represented as vertical gray lines in (b). Any surface S_i can be represented as a height function $f(x, y)$, which goes once through each column (x, y). In this case, the height function is $f(x, y) = x$. For the context of the MSF, the surface S_i can also be represented by attributing labels 0 (red) or 1 (yellow) for each node of N_i, as shown in (b). The yellow nodes with thicker outline refer to the nodes composing surface S_i. The next subsections will describe in more detail the structure of the created by the MSF, which includes N_i.

The minimum surface cost problem is a generalization for 3 dimensions of the minimum shortest path problem. The latter can be solved using the algorithm proposed by Dijkstra $et\ al.$ [65]. However, the minimum surface cost problem cannot be solved using this algorithm, but the MSF can formulate an equivalent problem in the context of the minimum closed set problem, which can be efficiently solved through minimum s-t cut/max-flow algorithms [95]. For that purpose, the MSF performs 4 operations [76] (Fig. 4.8):

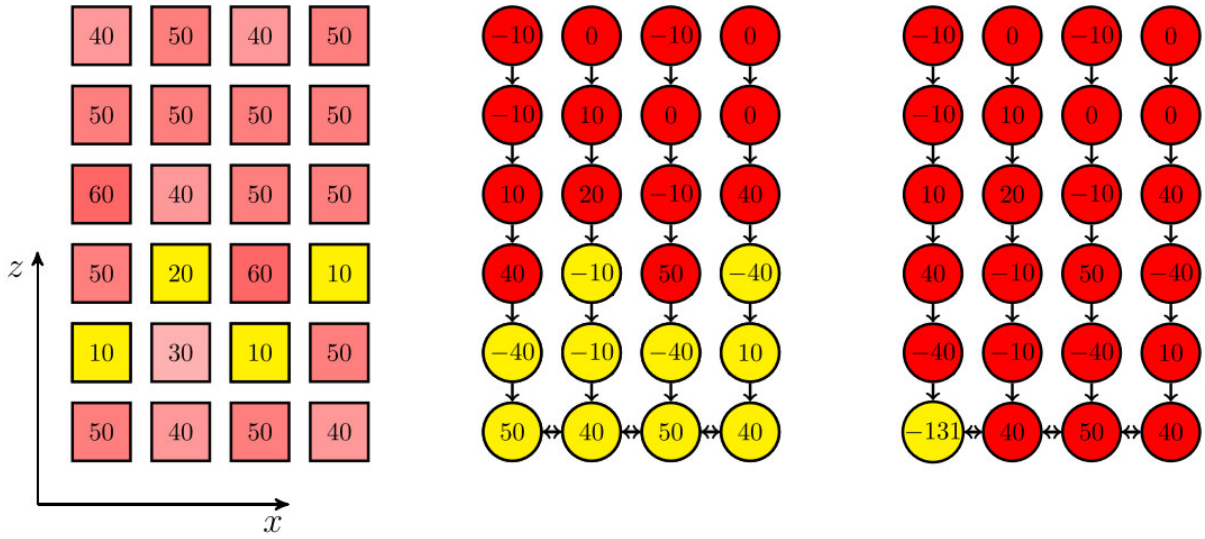
1. sets the nodes weights to $\phi_i(x, y, z)$;
2. adds intracolumn arcs;
3. adds intercolumn arcs for the bottom row;
4. adds a negative constant to a node of the bottom row.

After these operations, the minimum cost surface is retrieved as the top nodes of every (x, y) column from the minimum closed set of the graph G created by the MSF (Fig. 4.8).

The node weights of the minimum closed set problem ($\phi_i(x, y, z)$) are equivalent to those of the minimum surface cost problem ($\gamma_i(x, y, z)$) when $\phi_i(x, y, z)$ is defined as the discrete partial derivative of $\gamma_i(x, y, z)$ in the z direction (Fig. 4.8) [76]:

$$\phi_i(x, y, z) = \begin{cases} \gamma_i(x, y, z), & \text{if } z = 0 \\ \gamma_i(x, y, z) - \gamma_i(x, y, z-1), & \text{otherwise.} \end{cases} \tag{4.20}$$

Thus, $\gamma_i(x, y, \tau) = \sum_{z=0}^{z=\tau} \phi_i(x, y, z)$ for column (x, y) and height value τ (z direction).



(a) Minimum surface cost problem Weights γ_i; surface cost $= 50$.

(b) Minimum closed set problem Weights ϕ_i; set cost $= 50$.

(c) Adding a negative constant to ϕ_i of a bottom node of the graph in (b).

Figure 4.8: Correspondence between the minimum surface cost problem and the minimum closed set problem. The weights γ_i in (a) that are highlighted in yellow refer to the surface with minimum cost. The graphical model in (b) exhibits nodes with weights ϕ_i. The nodes highlighted in yellow have label 1 and represent the minimum closed set. The nodes in red have label 0. The top nodes of each column of the minimum closed set in (b) correspond to the minimum cost surface in (a). In (c), the sum of weights ϕ_i of the bottom row is always negative. For that, it was added to $\phi_i(x, y, 0)$ a constant of -181, resulting in $\phi_i(x, y, 0) = 50 - 181 = -131$ (highlighted in yellow in (c)). Thus, $\phi_i(x, y, 0)$ has a absolute value larger than the previous weight sum of the bottom row (i.e., 180 in image (b)). Furthermore, there are intercolumn arcs in the bottom row that impose that the nodes of this row must all have the same label. This ensures that the bottom row is always included in the minimum closed set. All depicted arcs have infinite capacities (images adapted from [95])

The intracolumn arcs added to graph G ensure that a surface only crosses a (x, y) column once. To fulfill that purpose, these arcs have infinite capacities and are added from $N_i(x, y, z)$ to $N_i(x, y, z - 1)$ (Fig. 4.8 (b)). This way, the intracolumn arcs establish that if a node belongs to the minimum closed set, every node below must also belong to it (Fig. 4.8 (b)).

The bottom row has intercolumn arcs and a negative constant added to one of its nodes to avoid having an empty set as solution. An empty set is a closed set, therefore it can also be a solution to the minimum closed set problem. The intercolumn arcs have infinite capacities (Fig. 4.8 (b) and (c)) to ensure that all nodes of the bottom row have the same label. The negative constant is added to $\phi_i(x, y, 0)$ (Fig. 4.8 (c)). The value of this constant is defined so the sum of bottom row weights ϕ_i is negative, thus assuring that the bottom row is always included[4].

[4]The optimization minimizes an energy function, therefore a bottom row with negative sum of weights is always included because it decreases the energy value.

4.1.2.2 Smoothness energy

The smoothness energy term $E_{smooth}(S_i)$ is composed by hard and soft constraints that regularize surface S_i. These will be defined in the following subsections for the x direction. For the y direction, the same approach can be followed.

4.1.2.2.1 Hard constraints

The hard smoothness constraints define the maximum and the minimum height variations[5] of a surface from one column to the next, which are represented for the x direction by $\Delta_x^{max}(x,y)$ and $\Delta_x^{min}(x,y)$, respectively. These limits are set for each pair of columns (x,y) and $(x+1,y)$, meaning that they can vary throughout the extension of the search space N_i. The maximum height variation limits ($\Delta_x^{max}(x,y)$) are translated by the MSF with the introduction of intercolumn arcs with infinite capacities to graph G. These arcs are set from $N_i(x+1,y,z+\Delta_x^{max}(x,y))$ to $N_i(x,y,z)$ (blue arrows in Fig. 4.9 (a)). The minimum height variation limits ($\Delta_x^{min}(x,y)$) are defined using a similar logic, .i.e., by adding arcs from $N_i(x,y,z)$ to $N_i(x+1,y,z-\Delta_x^{min}(x,y))$ (red arrows in Fig. 4.9 (a)). These arcs restrict the height variations of surfaces because any variation outside the defined limits $\Delta_x^{max}(x,y)$ and $\Delta_x^{min}(x,y)$ results in a set with outgoing arcs, i.e., sets that are not closed. The Fig. 4.9 (b) exemplifies a transition $\Delta f_i(x,y)=-1$ that is valid, since it is included in the limits $\Delta_x^{max}(x,y)=3$ and $\Delta_x^{min}(x,y)=-1$ (Fig. 4.9 (a)). For this transition, the set containing nodes with label 1 (depicted in cyan) is a closed set. On the contrary, the Fig. 4.9 (c) presents a case where the transition $\Delta f_i(x,y)=-2$ is outside the hard constraints limits ($\Delta_x^{min}(x,y)=-1$). In this situation, the set with nodes of label 1 (depicted in cyan) has one outgoing arc (highlighted in yellow), which prevents this transition from occurring.

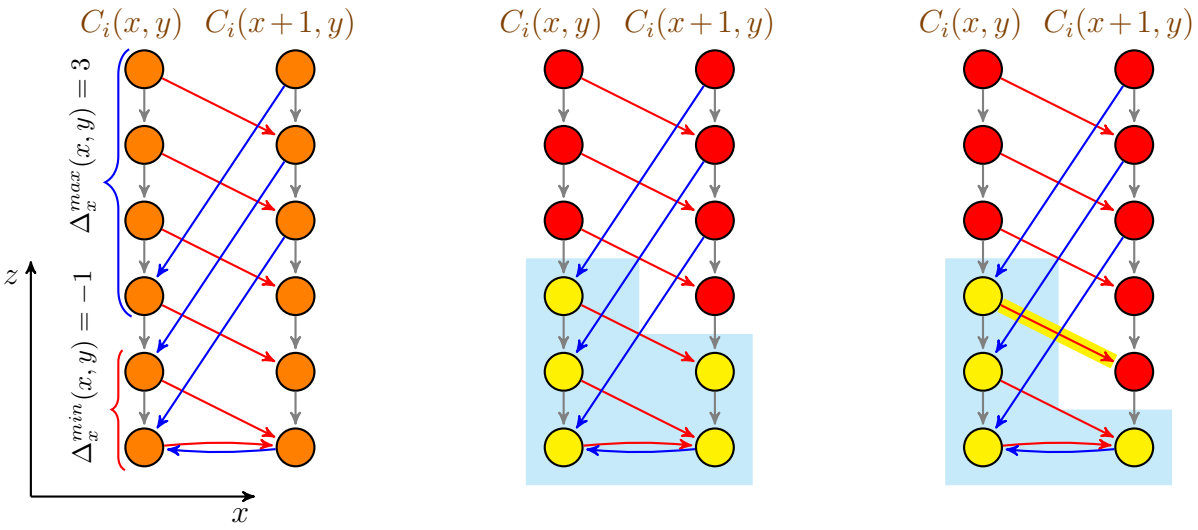
4.1.2.2.2 Soft constraints

The soft smoothness constraints impose a linear penalization to the height variations of surfaces for the x direction that deviate from a fixed value ($\Delta_x h^i_{fixed}$) [1]:

$$E_{smooth}(S_i)=\sum_{x\in X,y\in Y}\left(\lambda_{i_x}\left|\frac{\Delta f_i(x,y)}{\Delta x}-\Delta_x h^i_{fixed}(x,y)\right|\right), \tag{4.21}$$

where λ_{i_x} is the weight that controls the penalization and $\Delta f_i(x,y)/\Delta x$ is the height derivative of surface S_i in the x direction. To reflect these constraints in graph G, the MSF adds intercolumn arcs with a capacity of λ_{i_x} from $N_i(x,y,z)$ to $N_i(x+1,y,z+\Delta_x h^i_{fixed}(x,y))$ and from $N_i(x+1,y,z+\Delta_x h^i_{fixed}(x,y))$ to $N_i(x,y,z)$ (Fig. 4.10 (a)). By considering arcs with finite capacities, the MSF formulates a minimum s-excess problem (as referred in section 4.1.1). This means that the condition of closed set can be relaxed and outgoing arcs may exist, assuming that a penalty equal to the capacity of the arc is paid. This way, for every

[5] Height variations refers to variations in the z axis.



(a) Hard constraints

(b) $\Delta f_i(x,y) = -1$, valid transition

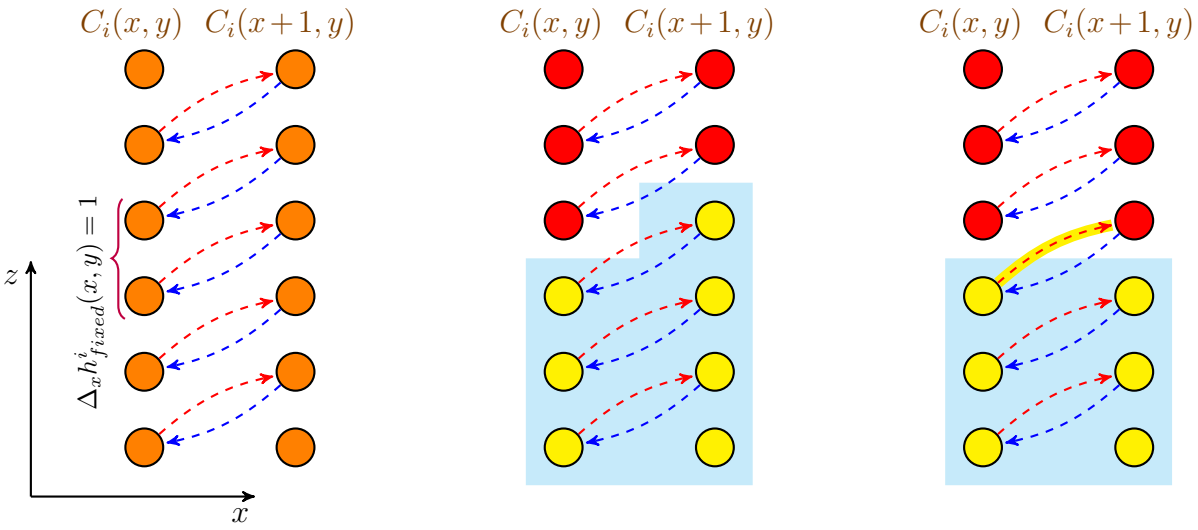
(c) $\Delta f_i(x,y) = -2$, invalid transition

Figure 4.9: Graphical model example of hard smoothness constraints. $C_i(x,y)$ refers to column (x,y) of the search space N_i of surface S_i. The solid gray arrows in correspond to the intracolumn arcs. The solid red and blue arrows refer to the intercolumn arcs that impose the hard constraints. All the arcs have an infinite capacity. The images (b) and (c) exemplify transitions between columns with height variations of $\Delta f_i(x,y) = -1$ and $\Delta f_i(x,y) = -2$, respectively. The sets that correspond to those transitions are depicted in cyan in (b) and (c). The yellow nodes within these regions have a label 1, while the red nodes have a label 0. In (b), the set in cyan is closed, thus the transition is valid. As for (c), the set in cyan has one outgoing arc (highlighted in yellow), implying that is not a closure and therefore this transition is not valid.

variation that differs in a single pixel above or below the $\Delta_x h^i_{fixed}$, the optimization needs to pay a penalty of λ_{i_x}, which results in the linear penalization defined in Eq. 4.21. Paying the penalty allows the nodes succeeding the outgoing arcs to be in the minimum s-excess set and have a label 0 (instead of label 1). This way, the soft smoothness constrains do not prohibit any transition, contrasting with the effect of the hard constraints. The Fig. 4.10 (b) exemplifies a case where the transition $\Delta f_i(x,y)$ is equal to $\Delta_x h^i_{fixed}$ (defined in Fig. 4.10 (a)). In this situation, the set with nodes of label 1 is a closed set, which implies that it is not necessary to pay any penalization for outgoing arcs. The Fig. 4.10 (c) on the other hand presents an example where $\Delta f_i(x,y)$ differs in one height unit from $\Delta_x h^i_{fixed}$ (defined in Fig. 4.10 (a)). In this case, the set containing nodes with label 1 has an outgoing arc (highlighted in yellow). Thus, for this transition to occur it is necessary to pay a cost of λ_{i_x} during the optimization.

4.1.2.3 Interaction energy

The interaction energy term $E_{inter}(S_i, S_j)$ is very similar to the smoothness energy term $E_{smooth}(S_i)$. The main difference is that the interactions are between subgraphs N_i and N_j instead of being contained just in



(a) Soft constraints

(b) $\Delta f_i(x,y) = 1$, valid transition

(c) $\Delta f_i(x,y) = 0$, valid transition if highlighted arc is cut

Figure 4.10: Graphical model example of soft smoothness constraints. $C_i(x,y)$ refers to column (x,y) of the search space N_i of surface S_i. The dashed red and blue arrows are represent the arcs that impose the soft constraints and their capacity is λ_{i_x}. The images (b) and (c) exemplify transitions between columns with height variations of $\Delta f_i(x,y) = 1$ and $\Delta f_i(x,y) = 0$, respectively. The sets that correspond to those transitions are depicted in cyan in (b) and (c). The yellow nodes within these regions have a label 1, while the red nodes have a label 0. In (b), the set in cyan is closed, thus it is not necessary to pay any penalty. As for (c), the set in cyan has one outgoing arc (highlighted in yellow), implying that it is required to pay a penalty of λ_{i_x}.

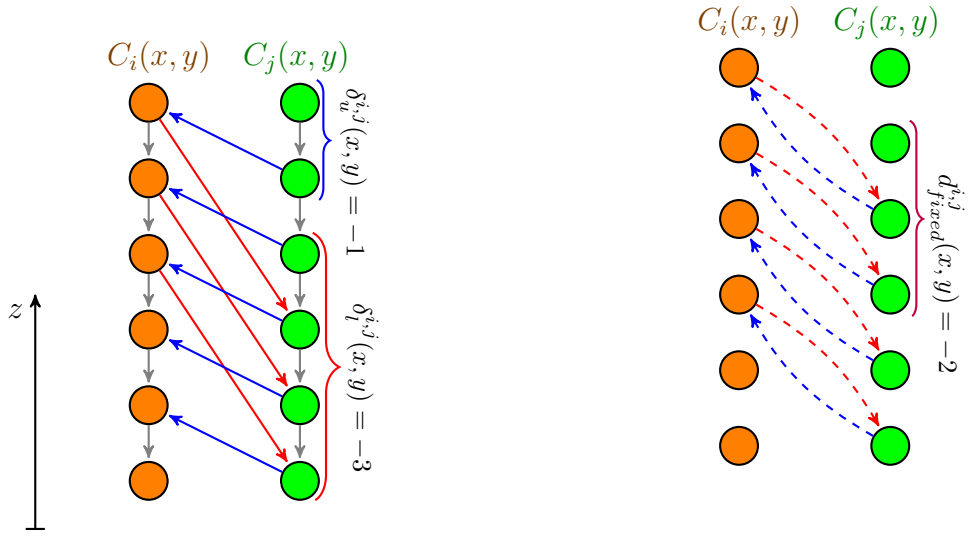
N_i. The interaction energy term is also composed by hard and soft constraints, which will be described in the following subsections for the x direction. The same logic can be applied for defining the interaction constraints for the y direction

4.1.2.3.1 Hard constraints

The hard constraints limit the distance between surfaces S_i and S_j. This is performed by setting the upper and lower limits of surface S_j in relation to surface S_i, which are represented by $\delta_u^{i,j}(x,y)$ and $\delta_l^{i,j}(x,y)$. The MSF defines the upper limit ($\delta_u^{i,j}(x,y)$) applying intergraph arcs with infinite capacities from $N_j(x,y,z-\delta_u^{i,j}(x,y))$ to $N_i(x,y,z)$ (blue arrows in Fig. 4.11 (a)). The lower limit ($\delta_l^{i,j}(x,y)$) is set by adding intergraph arcs from $N_i(x,y,z)$ to $N_j(x,y,z-\delta_l^{i,j}(x,y))$ (red arrows in Fig. 4.11 (a)).

4.1.2.3.2 Soft constraints

For each column (x,y), the soft interaction constraints penalize linearly the deviation of the distance between



(a) Hard constraints **(b)** Soft constraints

Figure 4.11: Graphical model example of hard and soft interaction constraints. The soft constraints of (b) impose a linear penalization. $C_i(x,y)$ and $C_j(x,y)$ respectively refer to columns (x,y) of the search spaces N_i and N_j of surfaces S_i and S_j, respectively. The orange and green nodes belong to N_i and N_j, respectively, which correspond to the search spaces of surfaces S_i and S_j. The solid gray arrows in (a) correspond to the intracolumn arcs with infinite capacity. The solid red and blue arrows of (a) refer to the intergraph arcs with infinite capacity that impose the hard constraints. The dashed red and blue arrows in (b) represent the arcs with capacity of $\alpha_{ij}/d_\sigma^{i,j}$, which impose the soft constraints.

surfaces S_i and S_j ($d^{i,j}$) in relation to a fixed distance value ($d_{fixed}^{i,j}$) [1]:

$$E_{inter}(S_i, S_j) = \sum_{x \in X, y \in Y} \epsilon_{ij} \left| d^{i,j}(x,y) - d_{fixed}^{i,j}(x,y) \right|, \tag{4.22}$$

where ϵ_{ij} is the weight that controls the linear penalization. The MSF applies these constraints by defining intergraph arcs with capacity ϵ_{ij} from $N_i(x,y,z)$ to $N_j(x,y,z+d_{fixed}^{i,j}(x,y))$ and from $N_j(x,y,z+d_{fixed}^{i,j}(x,y))$ to $N_i(x,y,z)$ (Fig. 4.11 (b)).

4.1.3 Local shape priors

In this work, the MSS problem is extended by integrating local shape priors to improve segmentation of the IRPEDC — one of the boundaries that limit drusen. Consequently, the MSF is also extended by formulating the local shape priors as SHOPs. To address high order potentials, first it is described how image segmentation can be posed as a MAP inference problem where the prior is modeled just with pairwise interactions. Then, high order potentials are addressed with particular focus on SHOPs.

4.1.3.1 MAP inference and B-K algorithm

Image segmentation can be posed as the estimation of the maximum $a\ posteriori$ (MAP) probability, which may be formulated as [96]:

$$\underset{\boldsymbol{Y}}{\text{argmax}} \qquad p(\boldsymbol{Y}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{Y})p(\boldsymbol{Y}), \tag{4.23}$$

where $\boldsymbol{X}$ is the observed data (e.g., image intensities) and $\boldsymbol{Y}$ are the corresponding class labels. This expression can be rewritten at the pixel level as:

$$\underset{\boldsymbol{Y}}{\text{argmax}} \qquad p(\boldsymbol{Y}|\boldsymbol{X}) \propto \prod_{i \in \boldsymbol{\zeta}} p(x_i|y_i)p(y_i), \tag{4.24}$$

where i refers to each pixel in the image domain ζ. To simplify the computational inference, this problem can be log-transformed and converted into a minimization:

$$\underset{\boldsymbol{Y}}{\text{argmin}} \qquad -\log p(\boldsymbol{Y}|\boldsymbol{X}) \propto -\sum_{i \in \boldsymbol{\zeta}} \log p(x_i|y_i) - \sum_{i \in \zeta} \log p(y_i). \tag{4.25}$$

There are several algorithms for solving MAP inference problems. Those algorithms typically solve a subset of MAP inference problems, which are frequently determined by several factors, such as: efficiency, number of labels and the model selected for the prior. In this work, the B-K algorithm was selected because it guarantees a global optimum solution and, at the same time, it is very efficient for problems related with computer vision [88]. Nevertheless, the algorithm restricts the problem to 2 labels (binary) and the prior term to being modeled by a MRF or CRF composed by submodular pairwise potentials. The pairwise potentials imply that only the relations between pairs of random variables are modeled. Furthermore, a pairwise potential is only submodular when the following condition is verified:

$$\kappa_p(y_i = 0, y_j = 0) + \kappa_p(y_i = 1, y_j = 1) \leq \kappa_p(y_i = 1, y_j = 0) + \kappa_p(y_i = 0, y_j = 1), \tag{4.26}$$

where κ_p refer to the pairwise potential, i and j are the random variables of the pairwise potential and their labels are represented by y_i and y_j, respectively. In practice, the submodularity condition means that the cost of the random variables i and j having the same label must be lower or equal to the cost of these random variables having different labels. Thus, pairwise potentials can be easily adapted to impose a smoothness prior[6] [97].

The MAP inference expression (Eq. 4.25) can be translated into the context of image segmentation using the B-K algorithm. This algorithm approaches MAP inference as min s-t cut problem, whose energy function is given by [88]:

$$E(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i \in \zeta} \kappa_u(x_i, y_i) + \sum_{i \in \zeta} \sum_{j \in \mathcal{N}_i} \kappa_p(y_i, y_j), \tag{4.27}$$

[6]The pairwise potentials of a submodular energy function cannot by applied for "contrastive" configurations when using the conventional max-flow algorithms, like the B-K algorithm [97].

where j are the pixels belonging to the neighborhood $\mathcal{N}_i$ of pixel i. The terms $\kappa_u(x_i, y_i)$ and $\kappa_p(y_i, y_j)$ refer respectively to the unary and pairwise potentials. The unary and pairwise potentials correspond respectively to the log-transformed likelihood ($-\log p(x_i|y_i)$) and prior ($-\log p(y_i)$) of Eq. 4.25. The energy function being minimized is equivalent to the log-transformed MAP value ($-\log p(\boldsymbol{Y}|\boldsymbol{X})$) of Eq. 4.25. In this case, the prior term is modeled by a MRF, since the pairwise potentials are not conditioned on observed data (x_i, x_j). If they were, then the prior would be modeled by a CRF instead.

Besides imposing submodularity, the B-K algorithm restricts the prior to be modeled just with pairwise interactions. This limits the expressiveness of the prior for capturing the complex image interactions [98, 99]. To try to solve this issue, some methods for optimizing functions with high order potentials have been proposed recently in the literature. In the next section, these type of potentials will be addressed, with particular focus on the class of SHOPs [100].

4.1.3.2 High order potentials

Higher order potentials have been used successfully for low level vision problems, such as image denoising, inpainting and texture restoration [101, 100]. Yet, the application in middle level problems (e.g. segmentation) is still limited due to the lack of efficiency in the optimization of higher order potentials [100]. Recently, several methods have been proposed to efficiently solve the energy functions with high order potentials. Some of those methods perform the optimization directly [98, 102, 103], while others transform the energy functions of higher order into a quadratic energy functions (order 2) [104, 99, 105]. One of the latter methods was proposed by Rother $et\ al.$ [100]. In this method, the energy function containing SHOPs is transformed into a quadratic non-submodular function. In this work, the MSF is extended by adding SHOPs. Therefore, the next section will be focused on this type of high order potentials and to their transformation into their pairwise counterparts.

4.1.3.3 Sparse high order potentials

A SHOP is simply a potential ψ applied to a clique c that favors a specific labeling L over all of the others (Fig. 4.12 (a)):

$$\psi_c(l_c) = \begin{cases} \theta_0, & \text{if } l_c = L \\ \theta_{max}, & otherwise, \end{cases} \tag{4.28}$$

where l_c is the labeling of clique c. The SHOP expression can also be presented in a more practical manner:

$$\psi_c(l_c) = \begin{cases} 0, & \text{if } l_c = L \\ \theta, & otherwise, \end{cases} \tag{4.29}$$

considering that $\theta = \theta_{max} - \theta_0$ and $\theta > 0$. By superimposing SHOPs, more than one labeling can be favored, however this has the downside of linearly increasing the number of auxiliary variables. In practical applications,

labelings l_c that differ slightly from L should have penalties similar to θ. Consequently, Rother $et\ al.$ [100] proposed a compact representation of SHOPs that uses only one auxiliary variable and penalizes the labelings l_c according to their Hamming distance from L. This results in a different type of potential: $\psi_c^g(l_c)$. This potential is supported by the deviation function g:

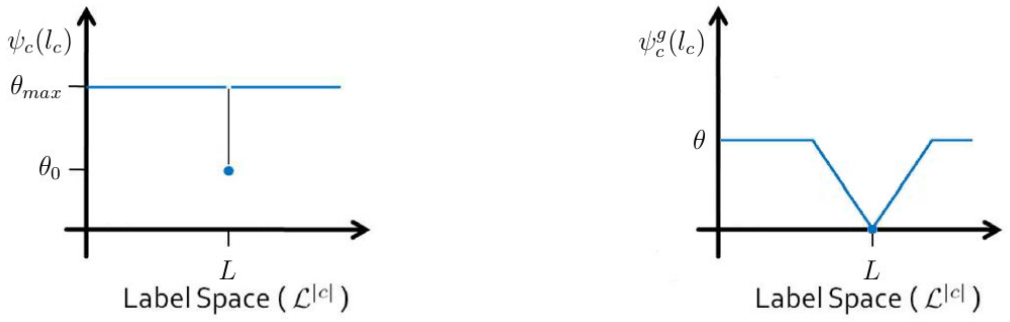
$$g(l_c) = \theta \sum_{v \in c} |w_v| \, \mathbb{1}_{l_c(v) \neq L(v)}, \tag{4.30}$$

$$\psi_c^g(l_c) = \min\{g(l_c), \theta\}, \tag{4.31}$$

where v is a vertex (or node) of clique c (i.e., $v \in c$), $l_c(v)$ is the label of vertex v and $L(v)$ is the label of the preferred labeling L for vertex v. The variable w_v controls the cost of $l_c(v)$ being different from $L(v)$[7]. The indicator function ($\mathbb{1}$) is 1 when $l_c(v) \neq L(v)$ and 0 otherwise. The deviation function g assigns a cost proportional to the Hamming distance between l_c and L (Fig 5.1). This function can be defined alternatively as:

$$g(l_c) = \theta \sum_{v|w_v \geq 0} w_v l_c(v) + \theta \sum_{v|w_v < 0} (-w_v)(1 - l_c(v)), \tag{4.32}$$

when the sign of weight w_v is defined in concordance with the label of the preferred labeling $L(v)$, i.e., $w_v < 0$ when $L(v) = 1$ and $w_v > 0$ when $L(v) = 0$.



(a) Original representation **(b)** Compact representation

Figure 4.12: Graph examples of SHOPs using different representations. The horizontal axis represents the label space, which contains every possible labeling of clique c. In (b), it is assumed that the labelings of the label space are ordered by Hamming distance from L, where labelings with smaller Hamming distances are closer to L and those with larger Hamming distances are farther from L (images adapted from [100]).

The new type of SHOP can be represented as $\psi_c^g(l_c) = \theta + \frac{\theta}{2} \min_{a \in \{0,1\}} F(a, l_c)$, where $F(a, l_c)$ corre-

[7]When this variable is constant, its value controls the slopes of the non-horizontal line segments of Fig. 4.12 (b).

sponds to the following quadratic pseudo-boolean function[8]:

$$F(a,l_c) = a(-2+\sum_{v\in c}|w_v|) + a\left(\sum_{v|w_v<0}(-w_v)(1-l_c(v)) - \sum_{v|w_v\geq 0} w_v(1-l_c(v))\right) + \\ + (1-a)\left(\sum_{v|w_v<0}(-w_v)l_c(v) - \sum_{v|w_v\geq 0} w_v l_c(v)\right) + \sum_{v|w_v\geq 0} w_v l_c(v) - \sum_{v|w_v<0}(-w_v)l_c(v), \tag{4.33}$$

where a represents the label of the auxiliary variable. Just as in Eq. 4.32, the sign of the w_c is related with $L(v)$, i.e., when $w_v < 0$ then $L(v) = 1$ and when $w_v > 0$ then $L(v) = 0$. The function F can be seen as 2 functions: one for $a = 0$ and another for $a = 1$. When $a = 0$ then the function F is equivalent to:

$$F(0,l_c) = \sum_{v|w_v<0}(-w_v)l_c(v) - \sum_{v|w_v<0}(-w_v)l_c(v) + \sum_{v|w_v\geq 0} w_v l_c(v) - \sum_{v|w_v\geq 0} w_v l_c(v) = 0, \tag{4.34}$$

and when $a = 1$ then the function F is given by:

$$F(1,l_c) = -2+\sum_{v\in c}|w_v| + \sum_{v|w_v\geq 0} w_v l_c(v) + \sum_{v|w_v<0}(-w_v)(1-l_c(v)) \\ - \sum_{v|w_v\geq 0} w_v(1-l_c(v)) - \sum_{v|w_v<0}(-w_v)l_c(v). \tag{4.35}$$

$F(0,l_c)$ is a constant, just like θ in Eq. 5.8, and $F(1,l_c)$ defines the penalizations for labelings l_c according to the Hamming distance (example in Table 4.1), just as the deviation function g in Eq. 5.8. The terms $\sum_{v|w_v\geq 0} w_v l_c(v) + \sum_{v|w_v<0}(-w_v)(1-l_c(v))$ from $F(1,l_c)$ are equal to function g without the scaling of θ (Eq. 4.32). Therefore, these terms apply a cost of w_v for each label $l_c(v)$ different from $L(v)$. The terms $-\sum_{v|w_v\geq 0} w_v(1-l_c(v)) - \sum_{v|w_v<0}(-w_v)l_c(v)$ also from $F(1,l_c)$ do the opposite, i.e., they apply a cost $-w_v$ for each label $l_c(v)$ equal to $L(v)$. Consequently, the cost difference between 2 arbitrary labelings differing in a single label is $2w_v$ (example in Table 4.1). This value is then adjusted to equal those of the potential $\psi_c^g(l_c)$ by the terms containing θ in $\psi_c^g(l_c) = \theta + \frac{\theta}{2}\min_{a\in\{0,1\}} F(a,l_c)$ (example in Table 4.1). The terms $-2+\sum_{v\in c}|w_v|$ of $F(1,l_c)$ are just constants used to set the value of potential $\psi_c^g(l_c)$ to 0 when $l_c = L$ (example in Table 4.1).

The quadratic pseudo-boolean function F can be represented by s-t graph (Fig. 4.13). The function F is quadratic because it only has terms with 1 or 2 variables, which can be translated into the graph as unary or pairwise potentials, respectively. For instance, the term $a\sum_{i\in c}(|w_i|-2)$ of F only has one variable: a. This means that this graph will represent this term as a unary potential, more concretely the unary potential represented by the arc that connects the nodes s and a (Fig. 4.13). Additionally, the term $a\sum_{v|w_v<0}(-w_v)(1-l_c(v))$ has several variables, but they interact only in pairs: a and $l(c)$. This term is translated as the pairwise

[8]The designation pseudo-boolean implies that the function domain is boolean ($\{0,1\}$), but is codomain is a real number ($\mathbb{R}$). The quadratic designation means that the function has order 2, i.e., only contains terms with 1 or 2 variables.

Table 4.1: Example of the values of the function F and the potential $\psi_c^g(l_c) = \theta + \frac{\theta}{2} \min_{a\in\{0,1\}} F(a, l_c)$ for all possible labelings l_c with 3 variables. The favored labeling is $L = (1, 1, 0)$. The labeling l_c matching the favored labeling L is highlighted in bold. The value of w_i is 1/2 when $L(v) = 0$ and -1/2 when $L(v) = 1$. The value of θ is 1. The expression HD (l_c, L) refers to the Hamming distance between labelings l_c and L.

$l_c(1)$	$l_c(2)$	$l_c(3)$	HD (l_c, L)	$F(0, l_c)$	$F(1, l_c)$	$\min_{a\in\{0,1\}} F(a, l_c)$	$\theta + \frac{\theta}{2} \min_{a\in\{0,1\}} F(a, l_c)$
1	**1**	**0**	0	0	-2	-2	0
0	1	0	1	0	-1	-1	1/2
1	0	0	1	0	-1	-1	1/2
1	1	1	1	0	-1	-1	1/2
0	0	0	2	0	0	0	1
0	1	1	2	0	0	0	1
1	0	1	2	0	0	0	1
0	0	1	3	0	1	0	1

potentials represented by the arcs connecting nodes v and a when $w_v < 0$ (i.e., when $L(v) = 1$). In Fig. 4.13, these arcs connect nodes 1 and 2 to a.

The function $F(a, l_c)$ is not submodular, meaning that the global optimum cannot be found by the B-K algorithm [106, 88]. Nevertheless, this function can be efficiently solved using the QPBO algorithm, which frequently finds the global optimum or at least a good approximation that contains part of the optimal solution [106].

Regardless of the dimension of the clique c, the SHOPs defined with the compact representation only have a single auxiliary node per potential, ensuring that the computational complexity of the problem is not overwhelming. This is an advantage of using SHOPs over other methods, which normally require the addition of at least a few auxiliary nodes per potential [100].

4.1.3.4 Extension of the Multiple Surface Segmentation problem

As mentioned earlier, the MSS problem is extended in this work by integrating local shape priors, which can be formally expressed by the minimization of the following energy function:

$$E(\boldsymbol{S}) = \sum_{i=1}^{n} (E_{bound}(S_i) + E_{smooth}(S_i)) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E_{inter}(S_i, S_j) + \sum_{i=1}^{n} \sum_{L\in\boldsymbol{L}_i} E_{SHOP}(S_i, L), \tag{4.36}$$

where the set of n surfaces is designated by $\boldsymbol{S}$ and the set $\boldsymbol{L}_i$ is composed by all favored labelings of the SHOPs used for surface S_i, being each of those favored labeling designated as L. As mentioned earlier, the MSF formulated the primordial terms of the MSS problem in the context of the minimum closed set or minimum s-excess problems. Afterwards, those problems were converted into a minimum s-t cut problem. In this case,

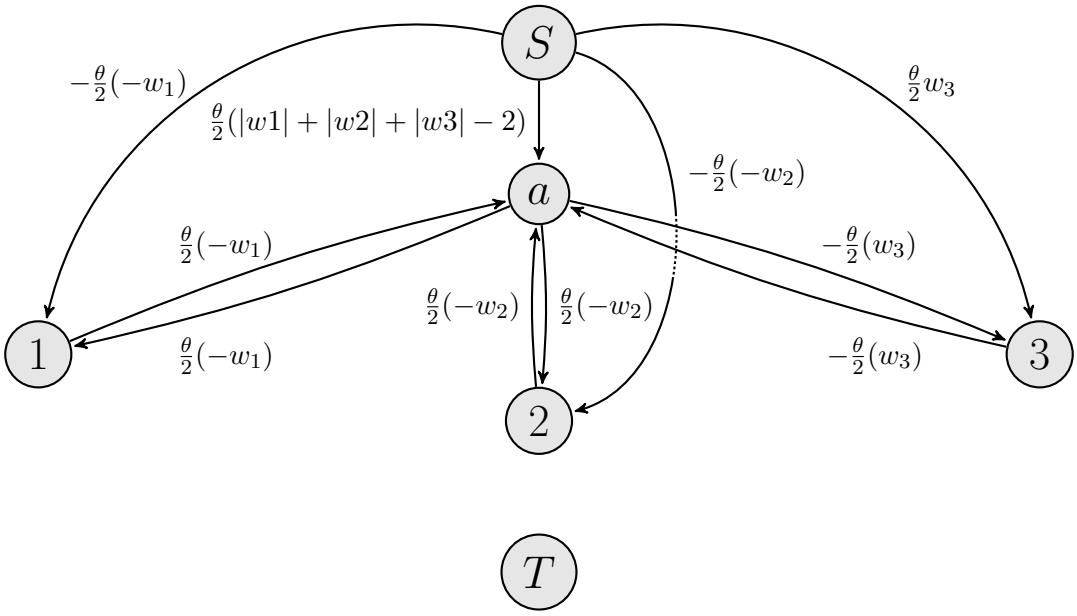


Figure 4.13: Graphical model of a higher order potential of type ψ_c^g with compact representation. In this example, the favored labeling is $L = (1, 1, 0)$ for nodes (1,2,3), respectively.

the SHOPs are already defined in the context of this problem, which implies that the MSF applies directly SHOPs when the MSS problem is converted for the context of the minimum s-t problem. Therefore, the term E_{SHOP} is closely related to potential ψ_c^g:

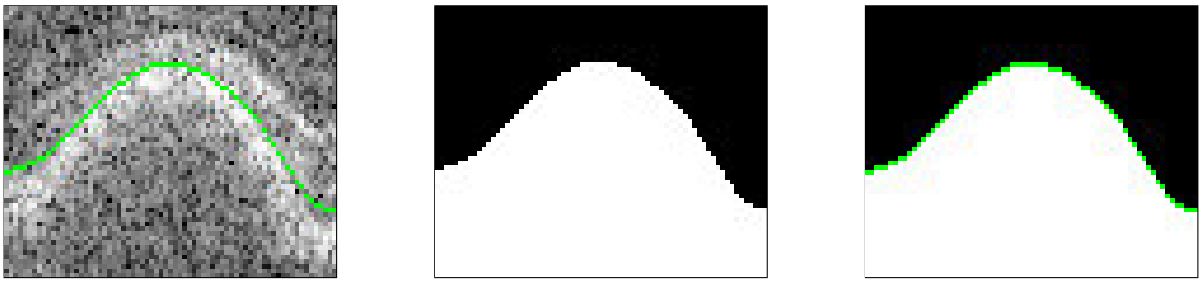
$$E_{SHOP}(S, L) = \psi_c^g(l_c, L), \quad \text{if } l_c = \Omega(S), \tag{4.37}$$

where function Ω converts surface S (height values) to a binary labeling (Fig. 4.14) and restricts it to the region where the ψ_c^g is applied, i.e. clique c. In the binary label, surface S corresponds to the top pixels with label 1, while pixels above and below the surface have labels 0 and 1, respectively (Fig. 4.14). Labelings can also be converted to surfaces using the inverse procedure, thus, applying a SHOP with a preferred labeling L means encouraging the segmentation to follow the surface encoded in L.

Those priors are formulated by the MSF as SHOPs. The SHOPs are defined in the context of the minimum s-t cut problem, instead of the minimum closed set or minimum s-excess problems as the other terms of the MSS problem.

4.2 Hybrid HMM-RF model

This section outlines the theoretical aspects necessary to comprehend the hybrid HMM-RF model. It starts by characterizing the RF classifier (subsection 4.2.1) and the HMM (subsection 4.2.2). Thereafter, the section presents how the RF can be integrated into a HMM to form the hybrid HMM-RF model (subsection 4.2.3).



(a) (b) (c)

Figure 4.14: Binary representation of the IRPEDC boundary. In (a) a drusen region is shown with respective IRPEDC segmentation depicted in green. The binary representation of IRPEDC segmentation is presented in (b), where black and white pixels represent pixels with labels 0 and 1, respectively. The IRPEDC corresponds to the top pixels with label 1 of each column of the binary labeling. In (c), the IRPEDC segmentation is overlaid in green over the binary labeling.

4.2.1 Random Forests

A RF is an ensemble classifier composed by decision trees [107]. As any other classifier, a RF finds a relation between the input feature vectors $\boldsymbol{v}$ and the target class labels c. In the training phase, each decision tree from the RF is grown by iteratively finding the best split of the training data (Fig. 4.15). A data split is performed by thresholding a particular feature. Thus, finding for the best split is translated in practice into searching for the combination of feature and threshold that produces the lowest value of Gini Impurity[9] [108]. With each data split, the decision trees grow by dividing the data by two nodes (Fig. 4.15). In the next iteration, the data of these nodes is divided again and the iterative process is repeated until the maximum depth or minimum number of samples is reached[10]. Additionally, the growing process can also halt when the nodes are pure – they only contain samples from one class.

Each of decision tree in the RF only considers a random subset of features and random subset of training data (i.e., bagging) [109]. Therefore, the predictions of each decision tree exhibit a low degree of correlation [109]. At test time, the RF outputs $a\ posteriori$ probability ($p(\mathrm{c}|\boldsymbol{v})$) that combines the information of all decision trees [109]:

$$p(\mathrm{c}|\boldsymbol{v}) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|\boldsymbol{v}), \tag{4.38}$$

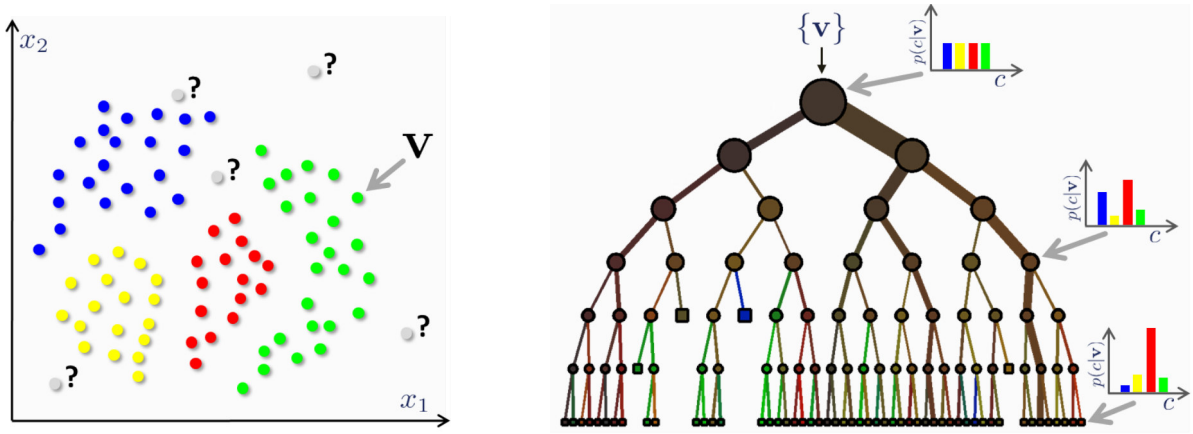
where T is the number of trees and $p_t(c|\boldsymbol{v})$ is the $a\ posteriori$ probability of each decision tree.

All of this grants the RF with a good generalization capability, as well as robustness to overfitting [109]. Besides these characteristics, the RF is also capable of multi-label classification without any additional alteration and it can handle efficiently large amounts of data, since it uses subsets of the training data and it is easily parallelizable[11] [109].

[9] Alternatively, it can be used information gain to find the best split of training data [108].

[10] The maximum depth and the minimum number of samples are defined by the user.

[11] It is easily parallelizable because the operations of each decision tree are computed independently.



(a) Data in feature space

(b) Decision tree training

Figure 4.15: Training a decision tree. Training data is depicted in (a) and a decision tree grown from that data is presented in (b). The classes of the training data are represented by the different colors and the points marked with a question marks are unseen data points. In (b), it is observable that the $a\ posteriori$ probability at deeper nodes is more discriminative than at nodes on the top of the tree. The amount of data that flows from a node to its children is presented by the width of each arc. Nodes at deeper levels of the tree contain less data due to the successive splits (image adapted from [109]).

The RF can also provide information regarding the importance of features. The importance of a feature is based on the mean Gini Impurity decrease for the whole forest caused by the splits using that particular feature [109]. With this information, unimportant features may be removed to reduce computational complexity. In some cases, the performance may even improve, since the removed features may contain noisy information.

4.2.2 Hidden Markov Model

A HMM is a statistical model used for representing probability distributions of sequences of observations [110]. This model assumes for sequence position k that an observation $\boldsymbol{O_k}$ was generated from a state H_k, which is not directly observable, i.e., it is hidden [110]. Furthermore, this hidden process satisfies the Markov property. For this case, the Markov property signifies that the value of state H_k only depends on the state value of the previous position in the sequence H_{k-1} [110]. Given these considerations, the joint probability of the sequence of observations $\boldsymbol{O}$ and hidden states $\boldsymbol{H}$ can be factored into the following expression (graphical representation presented in Fig. 4.16):

$$P(\boldsymbol{H},\boldsymbol{O}) = P(\boldsymbol{H})P(\boldsymbol{O}|\boldsymbol{H}) = p(H_1)p(\boldsymbol{O}_1|H_1)\prod_{k=2}^{K} p(H_k|H_{k-1})p(\boldsymbol{O}_k|H_k), \tag{4.39}$$

where K is the number of elements in the sequence. The factorization implies that the joint probability can be computed by determining the distributions of the initial state $p(H_1)$, of state transition $p(H_k|H_k-1)$ and of emission $p(\boldsymbol{O}_k|H_k)$. The HMM assumes a finite number of hidden states M. This implies that the initial state distribution is just a vector containing probabilities for the M states [110]. Moreover, the state transition

distribution is represented by a transition matrix with dimensions $M \times M$, which contains probabilities for every possible transitions between hidden states. The emission probability distributions reflect the probability of occurring an observation $\boldsymbol{O}_k$ generated by a hidden state H_k. These distributions are frequently modeled by Gaussian functions or by a Gaussian mixture models (GMM).

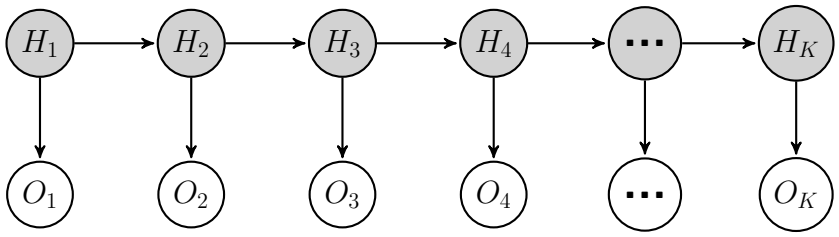


Figure 4.16: Bayesian network representing HMM. Shaded nodes correspond to hidden states and white nodes to observations.

Given the joint distribution $P(\boldsymbol{H}, \boldsymbol{O})$, it is possible to compute the most probable sequence of hidden states through the Viterbi algorithm [111]. For that purpose, this problem is converted into a problem of finding the shortest path in a graph [112]. The shortest path is then found by the Viterbi algorithm – a dynamic programming algorithm [112].

Other problems could be posed in the scope of HMM, as referred in [111], but these will not be addressed since they are not relevant under the scope of this thesis.

4.2.3 Combination of RF into HMM

In the literature of speech recognition, some works proposed different ways for integrating artificial neural networks (ANN) into HMM. Grezl $et\ al.$ [113] proposed using an ANN to extract features for a GMMs, which are used to represent the emission distributions[12]. The features are extracted from a neuron layer with a reduced number of neurons located at the middle of the ANN. These are designated as bottleneck features and have the advantage of combining all the features used in ANN into a reduced number of features, which may be controlled by the number of neurons defined for the middle layer. Before, employing the bottleneck features into GMM, the features are decorrelated through the a PCA or an equivalent method. Hermansky $et\ al.$ [114] proposed a similar approach that employs the outputs of an ANN as features of the GMM used for representing the emission distribution. The outputs of the ANN are not transformed by the softmax function and are decorrelated using a PCA. Renals $et\ al.$ [115] proposed a different approach, in which they argue that the emission distributions can be modeled with a ANN. This approach uses the ANN to model the $a\ posteriori$ distribution, which is then converted into a likelihood that models the emission distributions. In the next section, the focus will be in Renals $et\ al.$ [115] approach, since it will be applied for the hybrid HMM-RF model used for drusen detection.

[12]For the problem of speech recognition the emission distributions corresponds to the acoustic model [113].

4.2.3.1 Renals approach

The approach proposed by Renals $et\ al.$ [115] has been widely used with both conventional ANN [116, 117] and the more recent deep neural networks [118, 119, 120]. Modeling the emission distribution with a classifier has several benefits over conventional models such as the GMM. A classifier does not make hard assumptions regarding the statistical distribution of the observations [115]. Another advantage is the ease of introducing new features, such as context aware features. Furthermore, some classifiers, such as the RF, do not require the features to be decorrelated to be applied in practice, contrasting with the application of GMMs in some methods[13] [113, 114].

The implementation of the Renals $et\ al.$ [115] approach is straightforward, since a classifier can be integrated into a HMM just by converting the $a\ posteriori$ probability ($p(c|\boldsymbol{v})$) into likelihood ($p(\boldsymbol{v}|c)$) through the Bayes theorem:

$$p(\boldsymbol{v}|c) = \frac{p(c|\boldsymbol{v})p(\boldsymbol{v})}{p(c)}, \tag{4.40}$$

The likelihood $p(\boldsymbol{v}|c)$ corresponds to the emission distribution $p(\boldsymbol{O}|H)$ when the class label c is equivalent to hidden state H and vector $\boldsymbol{v}$ is equal to observation $\boldsymbol{O}$. This approach assumes that a classifier is capable of modeling the $a\ posteriori$ distribution $p(c|\boldsymbol{v})$. However, classifiers are only reported to accurately model this distribution when a few conditions are met. For instance, they must minimize metrics such as the mean squared error or cross entropy [122]. Finding the global minimum is not feasible for most problems, however it is reported in [115] that finding the minimum through cross-validation generally yields good results in practice. Correct estimation of the $a\ posteriori$ distribution also requires the classifier to be complex enough to model the distribution and that the train data is representative of the whole data of the problem [122].

To convert the $a\ posteriori$ distribution into a likelihood distribution, it is necessary to determine the prior ($p(c)$) and evidence ($p(\boldsymbol{v})$) distributions (Eq. 4.40). The prior distribution can be estimated from the training data based on the maximum likelihood criterion, i.e., by counting the frequencies of samples of each class and normalizing them. The evidence distribution is rather difficult to model [123]. In practical applications, this distribution is frequently considered to be uninformative, i.e., the evidence probability is set to be constant [123]:

$$p(\boldsymbol{v}|c) = \frac{p(c|\boldsymbol{v})}{p(c)} \times constant. \tag{4.41}$$

Regardless of the value of the constant, the Viterbi algorithm always provides the same results, since the constant value equally affects the cost of each possible sequence of hidden states [123].

The likelihood may be computed through alternative approaches, as suggested in [116, 117, 124]. Some applications obtain better results without dividing the $a\ posteriori$ probability by the prior term and applying it directly into the HMM [116, 117]. With this, it is possible to conclude that the underlying assumption of this approach is that the likelihood is just a scaled version of the $a\ posteriori$:

$$p(\boldsymbol{v}|c) = \eta\, p(\boldsymbol{v}|c), \tag{4.42}$$

[13] Some GMMs require the features to be decorrelated to avoid creating a very large model that would be impractical to use [121].

where η is a constant equal to the ratio $p(\boldsymbol{v})/p(c)$. For this ratio to be constant, both probabilities $p(\boldsymbol{v})$ and $p(c)$ have to be linearly correlated. This assumption could have some grounding, since the distribution $p(\boldsymbol{v})$ should be influenced by the relative proportions of classes in the data. Feature vectors characteristic of a particular class are expected to occur more frequently in that same class, meaning that the probability of occurrence feature vector of sample $p(\boldsymbol{v})$ from the majority class is possibly higher than for a sample of a minority class [117].

A particular case of this approach is when the data is balanced. In this case, $p(c) = 1/M$[14] and $p(\boldsymbol{v})$ is assumed to be constant:

$$p(\boldsymbol{v}|c) = p(c|\boldsymbol{v})_{balanced} \times \frac{1}{M} \times constant, \tag{4.43}$$

In this situation, considering the evidence probability as constant may be a reasonable assumption, since there is no influence of class imbalance. Thus, another approach for computing the likelihood is to perform a balanced training[15] of the RF and use the $a\ posteriori$ probability directly into the HMM [124].

All of these 3 approaches have been used to integrate a classifier into a HMM, however they basically differ on the assumptions make regarding the statistics of the data. For instance, some approaches assume that the prior distribution learned from training data ($p(c)$) is representative of the actual prior distribution. However, there is always some mismatch between training data and the actual prior distribution. Moreover, it may be assumed that the evidence distribution ($p(\boldsymbol{v})$) is uniform or that the distributions $p(\boldsymbol{v})$ and $p(c)$ are linearly correlated. Nevertheless, these assumptions may not be reasonable due to class imbalance or even due overlap of feature vectors from different classes. Therefore, the result of each approach is dependent of the data of a particular application. In the next chapter, these three approaches will be considered and one of them will be selected based on a few criteria.

4.3 Summary

This chapter started by reporting the theoretical background required to understand the extended MSS problem and the MSF that is used to solve it. First, the problems that support the MSF were described, as well as their conversions to the minimum s-t cut problem. Then, the MSS problem was described as the minimization of an energy function with 3 terms. The external boundary term attracts the surface to pronounced features of the image (e.g., large gradient magnitude values). The smoothness term limits the height variation of each surface and defines its rigidity. The interaction term limits the distance between surfaces and penalizes deviations from a fixed distance value. The MSF formulates each of these terms by creating a graph in the context of the minimum closed set or the minimum s-excess problems. Afterwards, the chapter moved to describing the local shape priors. These can be formulated in the form of SHOPs. In its original form, a SHOP is basically a potential where a labeling is favored above all others. A compact representation of SHOPs can generate a slightly different type

[14] M is the number of classes.

[15] A balanced training can be performed by methods such as subsampling the majority class and cost sensitive learning [108].

of potentials that penalize labelings according their Hamming distance relative to the favored labeling. Functions with this type of potentials can be converted to a non-submodular quadratic function, which can be efficiently solved by algorithms such as the QBPO. The MSF formulated every term of the MSS problem in the context of the minimum closed or minimum s-excess problems, which were then converted to a minimum s-t cut. The SHOPs are already formulated in the context of this problem. Therefore, the MSF applies these potentials only when the MSS problem is already converted into the context of the minimum s-t cut problem.

Thereafter, the chapter addressed the elements required for comprehending the hybrid HMM-RF model. First, the RF was addressed. This is an ensemble classifier composed by decision trees. This classifier has a good generalizing capability and is robust to overfitting. Moreover, it can efficiently handle large amounts of data because it is easily parallelizable. Then, the theoretical foundation of HMM was exposed. This is a statistical model used for representing probability distributions of sequences of observations. Given the initial state, transition state and emission distributions, the HMM can model the joint probability distribution of the sequence of observations and hidden states. Resorting to the Viterbi algorithm, the HMM can be used to find the most likely sequence of hidden states given a sequence of observations. A hybrid model can be created by integrating a classifier, such as an RF, into a HMM. For that purpose, it is required to convert the $a\ posteriori$ probability resulting from the classifier into a likelihood. To accurately model the $a\ posteriori$ distribution with a RF, it is necessary to meet some conditions like: finding the RF that attains the global minimum (or good estimate) in terms of mean squared error or cross-entropy, selecting a sufficiently complex RF to model the distribution and assuring that the training data is representative of the data of the whole problem. Integrating the RF into the HMM can be accomplished in different ways. Those present in the literature make some assumptions about the statistics of data, for instance the prior distribution is learned from the training data when there is always some mismatch to the whole data. Additionally, it may be assumed that the evidence distribution is uniform or that the ratio between the prior and evidence probabilities is constant. Thus, the results of each approach depend on the data of a particular application.

Chapter 5

Segmentation of Retinal Layers

In the previous chapter, the theoretical foundations of the extended MSF were described. This chapter will describe how this framework can be used to segment the limiting boundaries of drusen (i.e., the IRPEDC and the BM). In the subsection 5.1.1, the parameters of the extended MSF are set for the problem of segmenting retinal boundaries. Afterwards, in subsection 5.1.2, it is outlined an algorithm that segments the IRPEDC and the BM using the extended MSF. This algorithm was specifically developed to cope with the morphological changes caused by AMD, such as drusen and GA. After the description of the proposed method, it will be presented the experimental setup used for its evaluation (section 5.2). The setup characterizes several aspects of the evaluation such as: the database, implementations of other methods and the metric used for quantifying the results. Thereupon, in section 5.3, the proposed method is evaluated against the methods of Dufour $et\ al.$ [1] and of Chiu $et\ al.$ [85]. Additionally, the proposed method without SHOPs is also included in the evaluation to assess the effect of adding local shape priors.

5.1 Methods

5.1.1 Adaptation of the MSF to retinal boundary segmentation

In this subsection, the MSF parameters will be adapted to the segmentation of retinal boundaries in OCT images. Each energy term will be addressed in detail in the following subsections.

5.1.1.1 External boundary energy term

$E_{bound}(S_i)$ is the external boundary energy term that pushes each surface S_i to prominent features of the image. For that purpose, the values of $\gamma_i(x, y, z)$ must be defined as inversely proportional to the probability of a node $N_i(x, y, z)$ belonging to surface S_i, as previously mentioned in subsection 4.1.2.1. In the context of retinal boundary segmentation, the values of $\gamma_i(x, y, z)$ are often based on the vertical gradient of the image ($\nabla_z I(x, y, z)$). For boundaries corresponding to light-to-dark transitions ($\nabla_z I(x, y, z) < 0$), the values of γ_i

should be defined as $\gamma_i(x, y, z) = \nabla_z I(x, y, z)$, while for boundaries referring to dark-to-light transitions ($\nabla_z I(x, y, z) > 0$), the values of γ_i should be set as $\gamma_i(x, y, z) = -\nabla_z I(x, y, z)$.

In section 5.1.2.7, a segmentation of the IRPEDC is based on intensity information rather than on the gradient. The aim is to identify the surface with the highest intensity values. Thus, weights $\gamma_i(x, y, z)$ are defined as $\gamma_i(x, y, z) = I_{max}(y) - I(x, y, z)$. The intensity value is negated so the values of $\gamma_i(x, y, z)$ are inversely proportional to the probability of node $N_i(x, y, z)$ belonging to surface S_i. The constant $I_{max}(y)$ is optional, but it is still included because, in practice, it is simpler to keep the values of $\gamma_i(x, y, z)$ in the same range as the original intensity values.

5.1.1.2 Smoothness energy term

The hard smoothness constraints can be adapted to the segmentation of retinal boundaries by defining the limits of maximum and minimum height variations in the x direction, i.e., $\Delta_x^{max}(x, y)$ and $\Delta_x^{min}(x, y)$. For that, each column (x, y) of the height variation distribution in the x direction ($\Delta_x h^i(x, y)$) is modeled by a Gaussian function. By setting a confidence level for the Gaussian models, the values of $\Delta_x^{max}(x, y)$ and $\Delta_x^{min}(x, y)$ can be defined as follows [76]:

$$\Delta_x^{max}(x, y) = \Delta_x h_\mu^i(x, y) + \beta \cdot \Delta_x h_\sigma^i(x, y), \tag{5.1}$$

$$\Delta_x^{min}(x, y) = \Delta_x h_\mu^i(x, y) - \beta \cdot \Delta_x h_\sigma^i(x, y), \tag{5.2}$$

where β controls the confidence level of the Gaussian models. For instance, when $\beta = 2.6$, it is expected that about 99% of the values of $\Delta_x h^i(x, y)$ are within the confidence interval $[\Delta_x^{min}(x, y), \Delta_x^{max}(x, y)]$ [76]. Furthermore, $\Delta_x h_\mu^i(x, y)$ and $\Delta_x h_\sigma^i(x, y)$ refer respectively to the mean and standard deviation of the height variation distribution $\Delta_x h^i(x, y)$. The values of these variables can be computed from the prior dataset, thus adapting the height variation limits (i.e., Δ_x^{max} and Δ_x^{min}) to the problem of segmenting retinal boundaries [76]. The prior dataset is composed by height values $h^i(x, y)$ for each surface S_i retrieved from the manual segmentations of a training set.

Dufour et al. [1] set $\beta = 2.6$, which results in a confidence level around 99%. This is reasonable for healthy patients, however, for retinas with AMD, the 1% of the disregarded values are frequently related to abnormalities caused by large drusen. To avoid this issue, the confidence interval was widened by setting $\beta = 3.5$, which corresponds to a confidence level of 99.95%.

The soft smoothness constraints also need to be adapted to the segmentation of retinal boundaries. For that, it is necessary to define the values of $\Delta_x h_{fixed}^i(x, y)$ and λ_{i_x} (Eq. 4.21). As mentioned in subsection 4.1.2.2.2, the soft smoothness constraints penalize the height variations of surface S_i ($\Delta f_i(x, y)/\Delta x$) that differ from a fixed value $\Delta_x h_{fixed}^i(x, y)$. In this case, $\Delta_x h_{fixed}^i(x, y)$ was defined as $\Delta_x h_\mu^i(x, y)$, just as in [1]. This results in the following definition of the soft smoothness energy term:

$$E_{smooth}(S_i) = \sum_{x \in X, y \in Y} \left(\lambda_{i_x} \left| \frac{\Delta f_i(x, y)}{\Delta x} - \Delta_x h_\mu^i(x, y) \right| \right). \tag{5.3}$$

The variable λ_{i_x} is the weight that scales the penalization. The value of this variable was learned through a Bayesian optimization algorithm [125] in a 5-fold cross validation scheme[1].

5.1.1.3 Interaction energy term

The hard interaction constraints can also be adapted to the segmentation of retinal boundaries by defining the upper and lower limits of surface S_j in relation to surface S_i, which are represented as $\delta_u^{i,j}(x,y)$ and $\delta_l^{i,j}(x,y)$. For that, each column (x,y) of the distribution of distances between surfaces S_i and S_j ($d^{i,j}(x,y)$) is modeled by a Gaussian function. By setting a confidence level for the Gaussian models, the $\delta_u^{i,j}(x,y)$ and $\delta_l^{i,j}(x,y)$ can be defined as:

$$\delta_u^{i,j}(x,y) = d_\mu^{i,j}(x,y) - \beta \cdot d_\sigma^{i,j}(x,y), \tag{5.4}$$

$$\delta_l^{i,j}(x,y) = d_\mu^{i,j}(x,y) + \beta \cdot d_\sigma^{i,j}(x,y), \tag{5.5}$$

where $d_\mu^{i,j}(x,y)$ and $d_\sigma^{i,j}(x,y)$ refer, respectively, to the mean and standard deviation of the distribution of distances between surfaces S_i and S_j ($d^{i,j}(x,y)$). Analogous to the smoothness constraints, the smoothness constraints, the $\delta_u^{i,j}(x,y)$ and $\delta_l^{i,j}(x,y)$ can also be adapted to the segmentation of retinal boundaries using the information of the prior dataset to compute the values $d_\mu^{i,j}$ and $d_\sigma^{i,j}$.

The adaptation for the soft interaction constraints consists in defining the values of $d_{fixed}^{i,j}(x,y)$ and ϵ_{ij} (Eq. 4.22). As referred before in subsection 4.1.2.3.2, the soft interaction constraints penalize the deviation of the distance between surfaces S_i and S_j ($d^{i,j}(x,y)$) from a fixed distance value $d_{fixed}^{i,j}(x,y)$ and the variable ϵ_{ij} is the weight that scales the penalization. In this situation, the $d_{fixed}^{i,j}(x,y)$ was set to $d_\mu^{i,j}(x,y)$ and the variable ϵ_{ij} was defined as $\alpha_{ij}/d_\sigma^{i,j}(x,y)$, just as in [1]. With these alterations the interaction energy term is expressed as follows:

$$E_{inter}(S_i,S_j) = \sum_{x\in X, y\in Y} \frac{\alpha_{ij}}{d_\sigma^{i,j}(x,y)} \left| d^{i,j}(x,y) - d_\mu^{i,j}(x,y) \right|, \tag{5.6}$$

where α_{ij} is the weight controlling the penalization of deviating from the mean distance between surfaces S_i and S_j. The variable α_{ij} is learned through a Bayesian optimization algorithm [125] in a 5-fold cross validation scheme. By defining ϵ_{ij} as $\alpha_{ij}/d_\sigma^{i,j}(x,y)$, the relative importance of the soft interaction constraints is dependent of the variability of the training data. The purpose of dividing by $d_\sigma^{i,j}(x,y)$ is to reduce the penalty of deviations when these are more likely to occur (high variability) and increase the penalty for the opposing situation (low variability).

[1] More details on how the value of this variable was learned will be provided in subsection 5.2.3.

5.1.1.4 Local shape prior energy term

As mentioned in subsection 4.1.3.3, the compact representation of SHOPs $\psi_c^g(l_c)$ is defined as:

$$g(l_c) = \theta \sum_{v \in c} |w_v| \, \mathbb{1}_{l_c(v) \neq L(v)}, \tag{5.7}$$

$$\psi_c^g(l_c) = \min\{g(l_c), \theta\}, \tag{5.8}$$

where θ is the maximum cost of the potential and w_v refers to the cost of node v having a different label in relation to the favored labeling $L(v)$. This formulation is not practical for defining penalties for potentials of different sizes (or clique order $|c|$). For example, a potential value of a large clique could reach the maximum value (i.e., θ) quite easily just because of its dimension, since the potential is the sum of the costs of nodes having a label different from the favored labeling. Once the value θ is reached, the cost of the potential would be constant, implying that it would no longer be proportional to the Hamming distance. To avoid this issue, it is proposed a different formulation of the parameters θ and w_v:

$$\theta = p|c| \;\; , \;\; w_v = \begin{cases} 1/|c|, & \text{if } \; L(v) = 0 \\ -1/|c|, & \text{if } \; L(v) = 1. \end{cases} \tag{5.9}$$

With this alteration, the SHOP always exhibits a linear penalization:

$$\psi_c^g(l_c, L) = \sum_{v \in c} p \, \mathbb{1}_{l_c(v) \neq L(v)}, \tag{5.10}$$

where p is the penalty increment for each element $l_c(v)$ that differs from the favored labeling $L(v)$ (see example in Fig. 5.1). With this formulation, the potential ψ_c^g is never truncated at θ, since this variable is set as the maximum sum of w_v, i.e., the cost of having a labeling l_c being completely different from the favored labeling L (example in Fig. 5.1 (d)). The value of p is learned through a Bayesian optimization algorithm [125] in a 5-fold cross validation scheme[2].

(a) Favored labeling L p = 5. **(b)** HD = 1, ψ_c^g = 5. **(c)** HD = 3, ψ_c^g = 15. **(d)** HD = 9, ψ_c^g = 45.

Figure 5.1: Example of a SHOP with compact representation. The favored labeling L is presented in (a). The Hamming distances (HDs) are relative to (a) and the values of ψ_c^g from segmentations (b), (c), (d) consider that p = 5, leading to θ = 45. In this situation, the Hamming distances are equal to the number of pixels with labels different from those of the favored labeling L.

[2]More details on how the value of this variable was learned will be provided in subsection 5.2.3.

5.1.2 Proposed segmentation algorithm

In this thesis, it is proposed an algorithm that segments the IRPEDC and the BM even in the presence of lesions related to AMD (Fig. 5.2), such as drusen and GA. The ILM and the IS-OS are also segmented (Fig. 5.2), but these are only used as auxiliary boundaries. The segmentation result of each step of the algorithm is used to restrict the search spaces of the following ones, thus reducing the computational complexity and avoiding undesired interactions between boundaries. An overview of the proposed segmentation algorithm is presented in Fig. 5.3. Each step of the algorithm will be described in more detail in the next subsections.

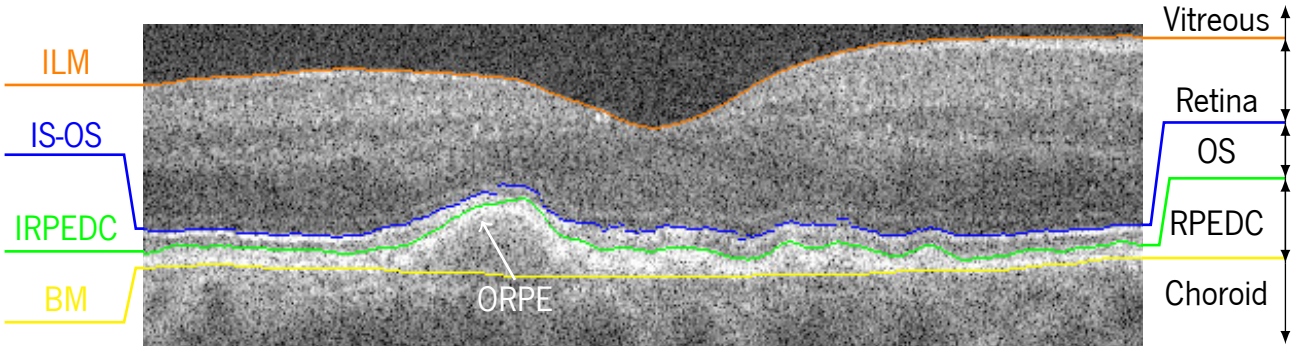


Figure 5.2: Example of the segmentation results of the proposed method.

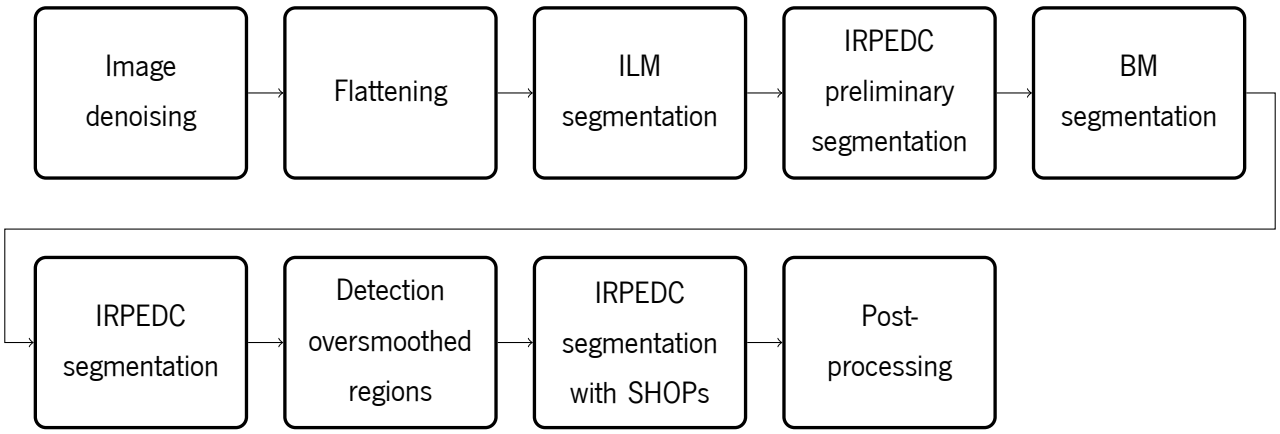


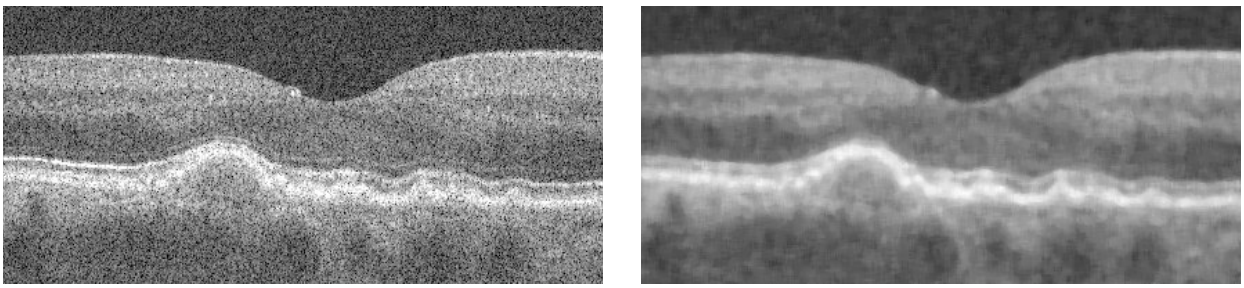
Figure 5.3: Diagram of the algorithm proposed for segmenting the IRPEDC and the BM.

5.1.2.1 Step 1: Image denoising

To attenuate the speckle noise, a median filter was applied to each B-scan (Fig. 5.4). This filter was used because it preserves edges and it is robust to multiplicative noises [126, 127], such as speckle noise. Additionally, the median filter has low computational complexity, resulting in a relatively fast denoising.

In a few cases, a Gaussian filter was used instead of the median filter. This filter can also be employed for attenuating speckle noise. However, the main purpose of using the Gaussian filter was to smooth edges

containing high frequencies and try to retain other edges composed by lower frequencies. The exceptional cases in which the Gaussian filter is used will be explicitly mentioned in the following subsections.

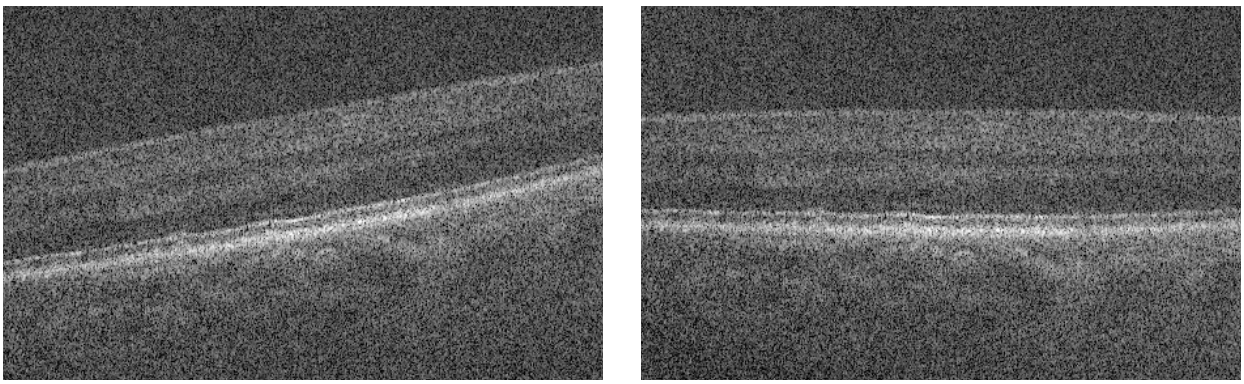


(a) Original image **(b)** Denoised image

Figure 5.4: Denoising of an OCT B-scan with a median filter (step 1).

5.1.2.2 Step 2: Flattening

In this step, a preliminary joint segmentation of the ILM and the IRPEDC is performed. This segmentation only used hard constraints to reduce computational complexity. The hard constraints (Eqs. 5.1, 5.2, 5.4 and 5.5) assume a Gaussian distribution for each slice (or y position), implying that all columns within a B-scan will have the same hard constraints limits. These constraints are set in this way because the position of the fovea is not known yet, thus the manual segmentations cannot be aligned to create a meaningful prior dataset for each x position (B-scan column). After the joint segmentation, a 2^{nd} order polynomial fit of the segmentation of the IRPEDC is used to flatten each slice (Fig. 5.5). The flattening is accomplished by moving columns up and down until the fitted boundary is horizontal.



(a) Original image **(b)** Flattened image

Figure 5.5: Flattening of an OCT B-scan (step 2).

5.1.2.3 Step 3: ILM segmentation

After the flattening, the fovea is detected as the lowest point of the smoothed preliminary ILM segmentation (step 2) from the central B-scan. This information is then used to align the manual segmentations and recalculate the prior dataset.

Dufour $et\ al.$ [1], defined the search space of most segmentations using a previously segmented boundary and the hard interaction constraints limits ($\delta_u^{i,j}$ and $\delta_l^{i,j}$ from Eqs. 5.4 and 5.5), which refer to the maximum and minimum distances between any two boundaries. Here, the same strategy is used (Fig. 5.6). The reference for computing the search space of the ILM boundary is the flattened version of the IRPEDC from the previous step. This segmentation might include segments of the IS-OS, since this boundary and the IRPEDC are spatially close and the preliminary segmentation of the IRPEDC from step 2 did not include the IS-OS. To handle this issue, the top limit of the search space of the ILM is calculated using maximum distance between the IRPEDC and the ILM ($\delta_l^{IRPEDC,ILM}$), while the bottom limit uses the minimum distance between the IS-OS and the ILM ($\delta_u^{IS-OS,ILM}$), as depicted in Fig. 5.6.

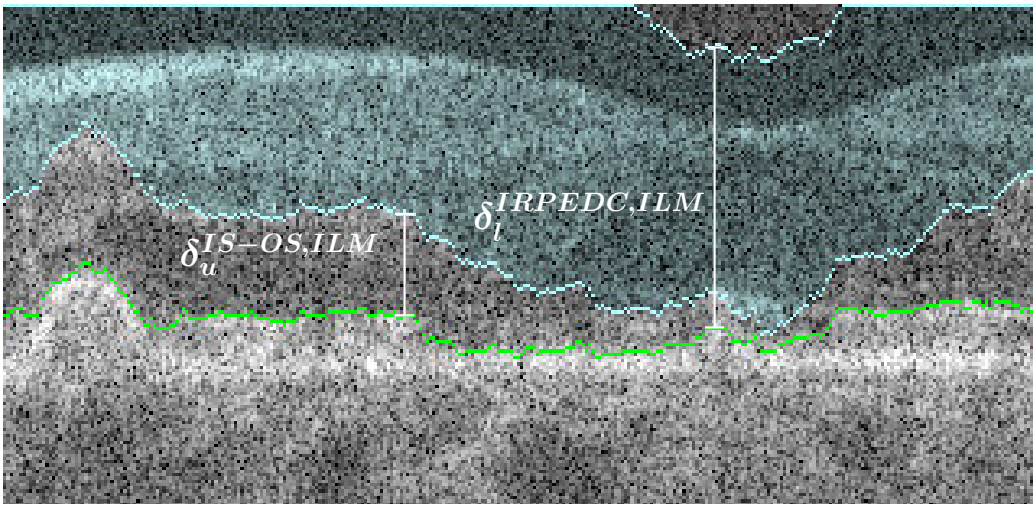


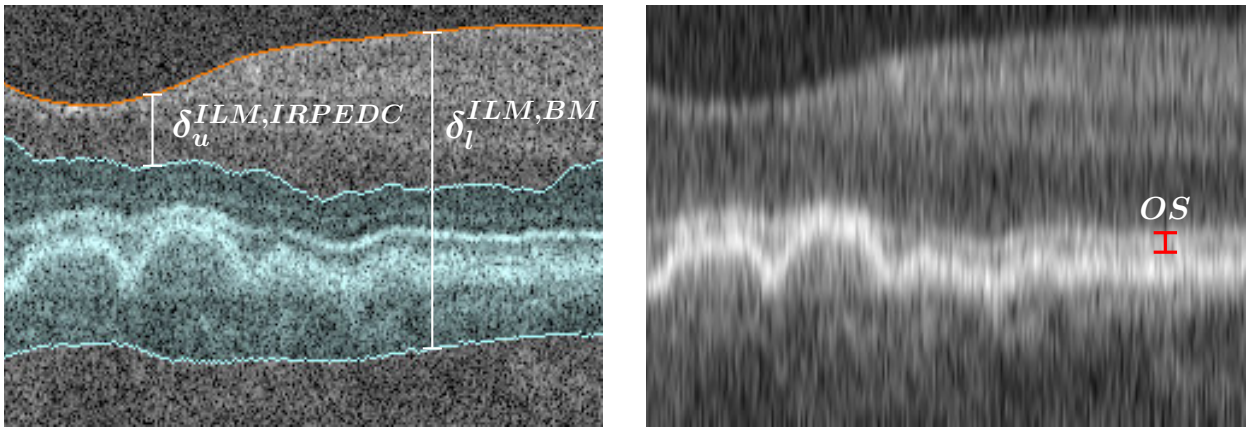
Figure 5.6: Search space used in step 3 for segmenting the ILM boundary (cyan). The preliminary segmentation of the IRPEDC from step 2 is presented in green.

5.1.2.4 Step 4: IRPEDC preliminary segmentation

In this step, a joint segmentation of the IRPEDC and the BM is performed with the aim of acquiring a preliminary segmentation of the IRPEDC. This preliminary segmentation will be used for defining the top limit of the search space for segmenting the BM individually in the next step.

The search space for the joint segmentation is determined by using the ILM as reference in conjunction with the minimum distance between the ILM and the IRPEDC ($\delta_u^{ILM,IRPEDC}$) and the maximum distance between the ILM and the BM ($\delta_l^{ILM,BM}$), as shown in Fig. 5.7 (a).

To avoid interference from the sharp dark-to-light transitions of the IS-OS and the light-to-dark transitions from inside the OS, the segmentation is carried out in an image obtained by applying a 1D Gaussian filter in the vertical direction (Fig. 5.7 (b)). The sharp transitions are mainly composed by high frequencies and the Gaussian filter acts as a low-pass filter. Therefore, these type of transitions are considerably more smoothed than transitions composed by lower frequencies, such as those presented in the IRPEDC and the BM.



(a) Search space of step 4 **(b)** Gaussian filter denoising

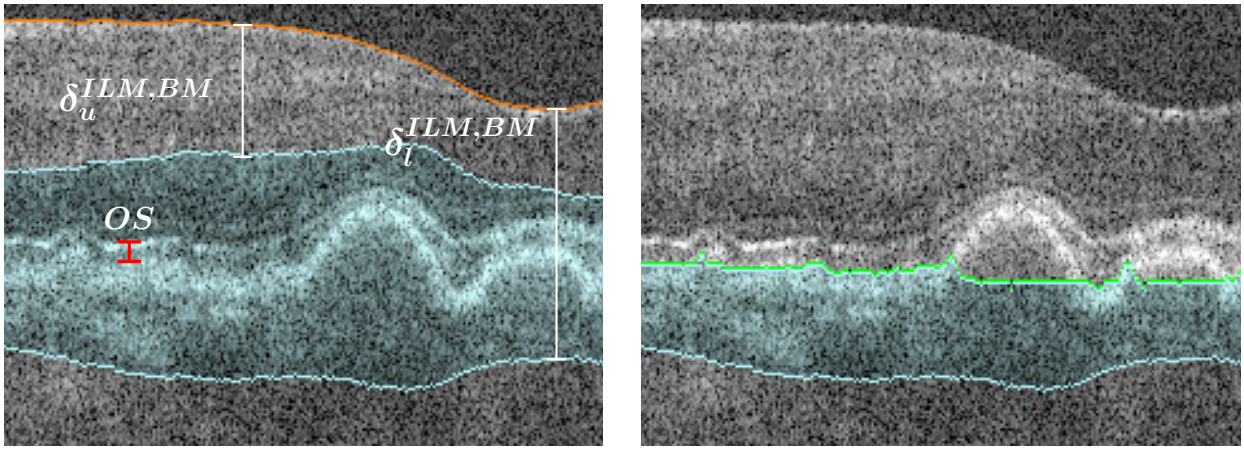
Figure 5.7: Search space used in step 4 for the preliminary joint segmentation of the IRPEDC and the BM boundaries (cyan) and the denoising using a vertical Gaussian filter. The ILM segmentation from step 3 is presented in orange in (a). The OS is marked in red in (b).

5.1.2.5 Step 5: BM segmentation

Following the same strategy as in step 3, the search space for the individual segmentation of the BM could be obtained by using the ILM as reference and the hard interaction constraints limits between the ILM and the BM ($\delta_u^{ILM,BM}$ and $\delta_l^{ILM,BM}$), as exemplified in Fig. 5.8 (a). However, this search space would often include sharp light-to-dark transitions from within the OS (Fig. 5.8 (a)), which would interfere with the segmentation of the BM. To exclude such transitions, the previous step performed a preliminary segmentation of the IRPEDC. With this alteration, the BM can be segmented individually by altering the top limit as the IRPEDC segmentation of step 4 with a unitary downward offset (Fig. 5.8 (b)). The bottom limit remains the same, i.e., the ILM is used as reference along with the value of $\delta_l^{ILM,BM}$ (Fig. 5.8 (a)).

5.1.2.6 Step 6: IRPEDC segmentation

In this step, the IRPEDC and the IS-OS are segmented in conjunction. This is necessary because these two boundaries are spatially close and present similar dark-to-light transitions. If the IRPEDC was not segmented in conjunction with the IS-OS, it could be easily attracted to the IS-OS. The search space for the joint segmentation of the IRPEDC and the IS-OS is defined by two references: the ILM and the BM, for the upper and lower limits,



(a) **(b)**

Figure 5.8: Two search spaces that could be used in step 5 for segmenting the BM boundary (cyan). In (a), the search space includes the OS (marked in red). This layer contains light-to-dark transitions that could hinder the segmentation of the BM. Therefore, it was performed a preliminary segmentation of the IRPEDC in step 4 to exclude this transition from the search space as shown in (b). Immediately above the search space of (b), it is depicted the preliminary segmentation of the IRPEDC from step 4 in green.

respectively. Thus, the hard constraints limits used are $\delta_u^{ILM,IS-OS}$ and $\delta_l^{BM,IRPEDC}$ (Fig. 5.9). This strategy is the same as the one used by Dufour $et\ al.$ [1] for segmenting boundaries between the ILM and the IS-OS.

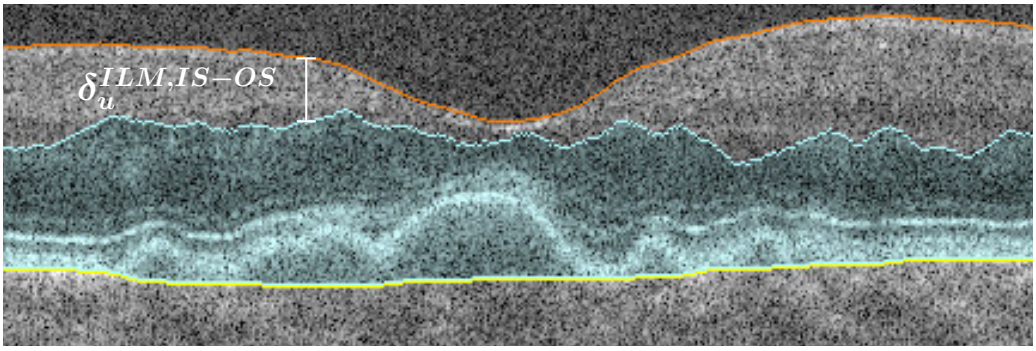


Figure 5.9: Search space used in step 6 for the joint segmentation of the IS-OS and the IRPEDC boundaries (cyan). The ILM is presented in orange and the BM corresponds to the yellow boundary. For the whole image, $\delta_l^{BM,IRPEDC} = 1$, thus it was not depicted in this image.

5.1.2.7 Step 7: Detection of oversmoothed regions

In this step, the IRPEDC boundary is resegmented without soft smoothness constraints and individually, i.e., without interaction constraints of the IS-OS. This segmentation is less restricted in order to properly segment regions where the OS is thin or there are large local height variations of the IRPEDC, such as drusen regions[3]

[3]This was inspired by the detection of pigment epithelial detachments in [50].

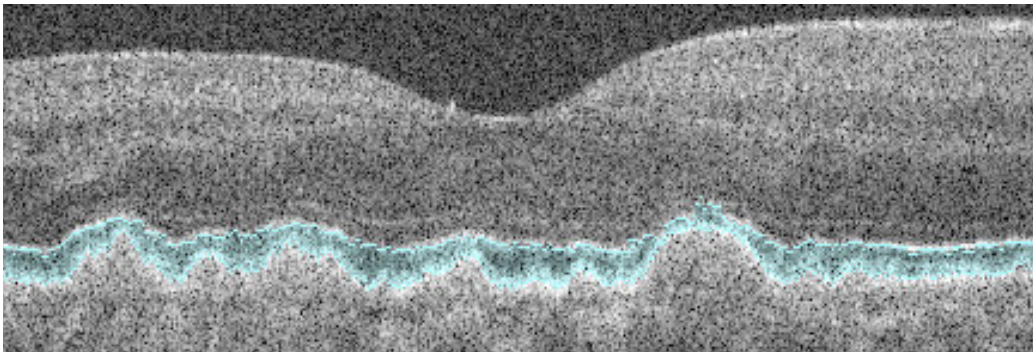
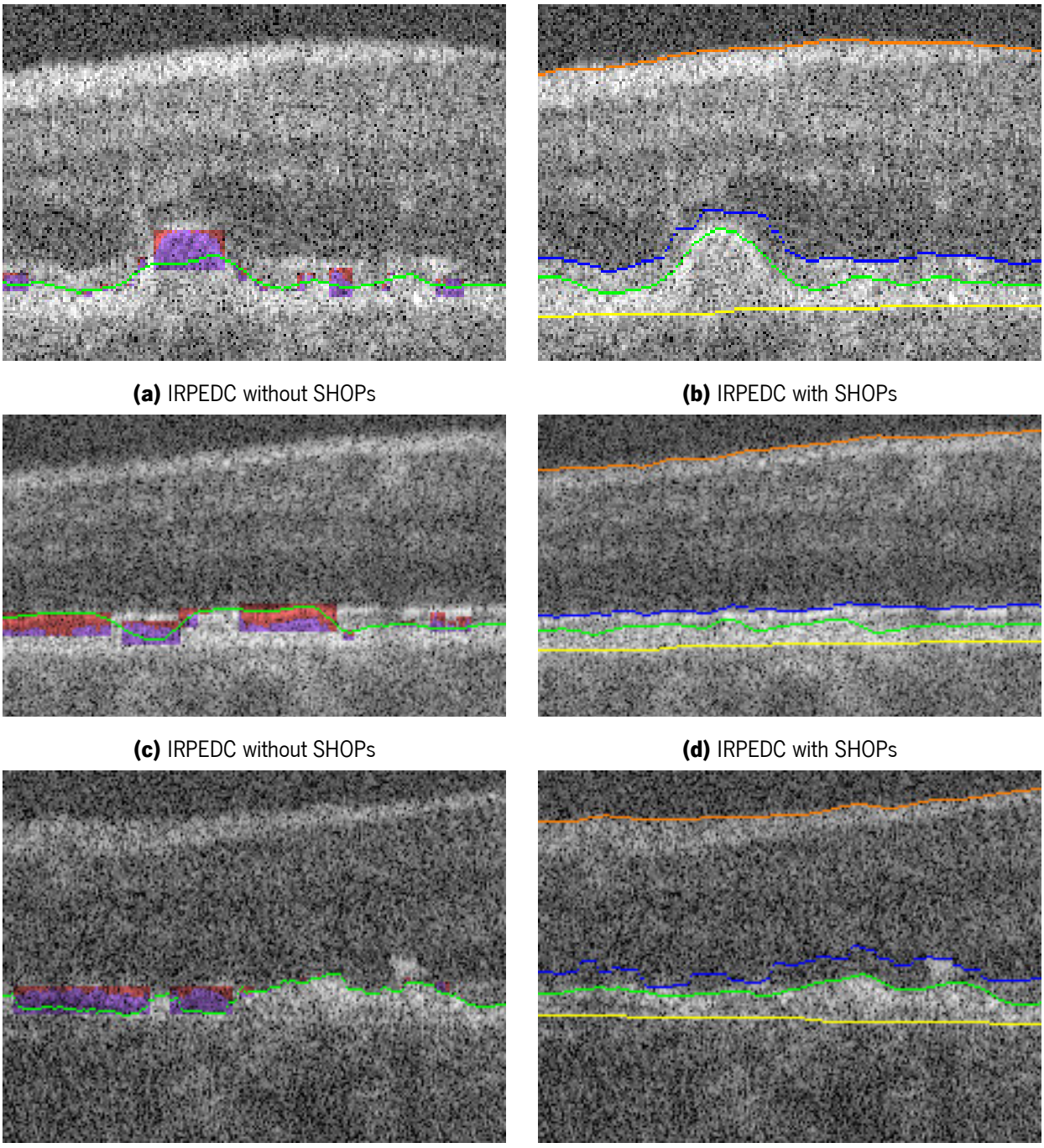


Figure 5.10: Search space used in step 7 for the resegmentation of the IRPEDC boundary (cyan). The upper bound is the IS-OS with an offset for each column that depends on the intensity gradient. The lower bound is composed by the brightest pixels of a Gaussian filtered image.

(Fig. 5.11 (a) and (b)) [13, 1]. The less restricted and individual segmentation of the IRPEDC can be attracted to the IS-OS or even to dark-to-light transitions within the RPEDC layer. These issues can even happen when the IRPEDC is segmented in conjunction with the IS-OS (Fig. 5.11 (c) and (e)), as performed in previous step. To avoid these issues, the search space of the individual IRPEDC segmentation has to be carefully defined. The upper bound is defined by using the IS-OS with a slight offset. The offset is determined for each column by using the intensity derivative to search for an extremum a few pixels below the IS-OS (Fig. 5.10). The extremum will correspond to either the maximum of the white band of the OS or the minimum of the dark band of the OS. Either way, the top limit will be below the IS-OS and above the IRPEDC. For the lower bound, the brightest pixels of the RPEDC are segmented by modifying the weights of the E_{bound} term (section 5.1.1.1) to be dependent on the intensity of the image after smoothing with a 1D Gaussian filter in the vertical direction, similar to what was done in [85]. This, segmentation did not employ soft constraints to avoid oversmoothing of drusen. The resulting segmentation corresponds approximately to the middle of the RPEDC layer, which can be used as the lower bound for the search space (Fig. 5.10). Using this lower bound excludes some dark-to-light transitions inside the RPEDC that could hinder the segmentation (Fig. 5.11 (e)). Those transitions are related to with noise or morphological changes caused by AMD. Moreover, the segmentation of the lower bound uses intensity information which makes it more robust to regions where the IRPEDC exhibits low contrast or to regions with foci of hyperreflective material inside drusen.

The difference between the IRPEDC segmentations from this step and the previous produces a binary image (Fig. 5.12 (b) and (c)). This image is composed by several objects[4] whose limits are used to define a rectangular neighborhood where each SHOP is applied, i.e. the clique c (rectangles in cyan of Fig. 5.12 (c)). Afterwards, the SHOPs are defined by converting the IRPEDC segmentation of this step to a labeling and restricting it to the rectangular neighborhoods (Fig. 5.12 (d)).

[4]The objects are defined according to a connectivity of 8 pixels.



(a) IRPEDC without SHOPs

(b) IRPEDC with SHOPs

(c) IRPEDC without SHOPs

(d) IRPEDC with SHOPs

(e) IRPEDC without SHOPs

(f) IRPEDC with SHOPs

Figure 5.11: Situations in which the SHOPs can improve the segmentation of the IRPEDC (green). The first row refers to the situation where a drusen is oversmoothed, the second row to the case of the IRPEDC segmentation being attracted to the IS-OS and the third row to the attraction of the IRPEDC to transitions within the RPEDC. The images in the left column show the IRPEDC without SHOPs (step 6) and the favored labelings of SHOPs overlaid in red and purple. The red and purple pixels represent respectively the labels 0 and 1 of the favored labelings. In the right column, the segmentations of the IRPEDC already use the SHOPs (step 8). The remaining depicted boundaries are: ILM (orange), IS-OS (blue) and BM (yellow).

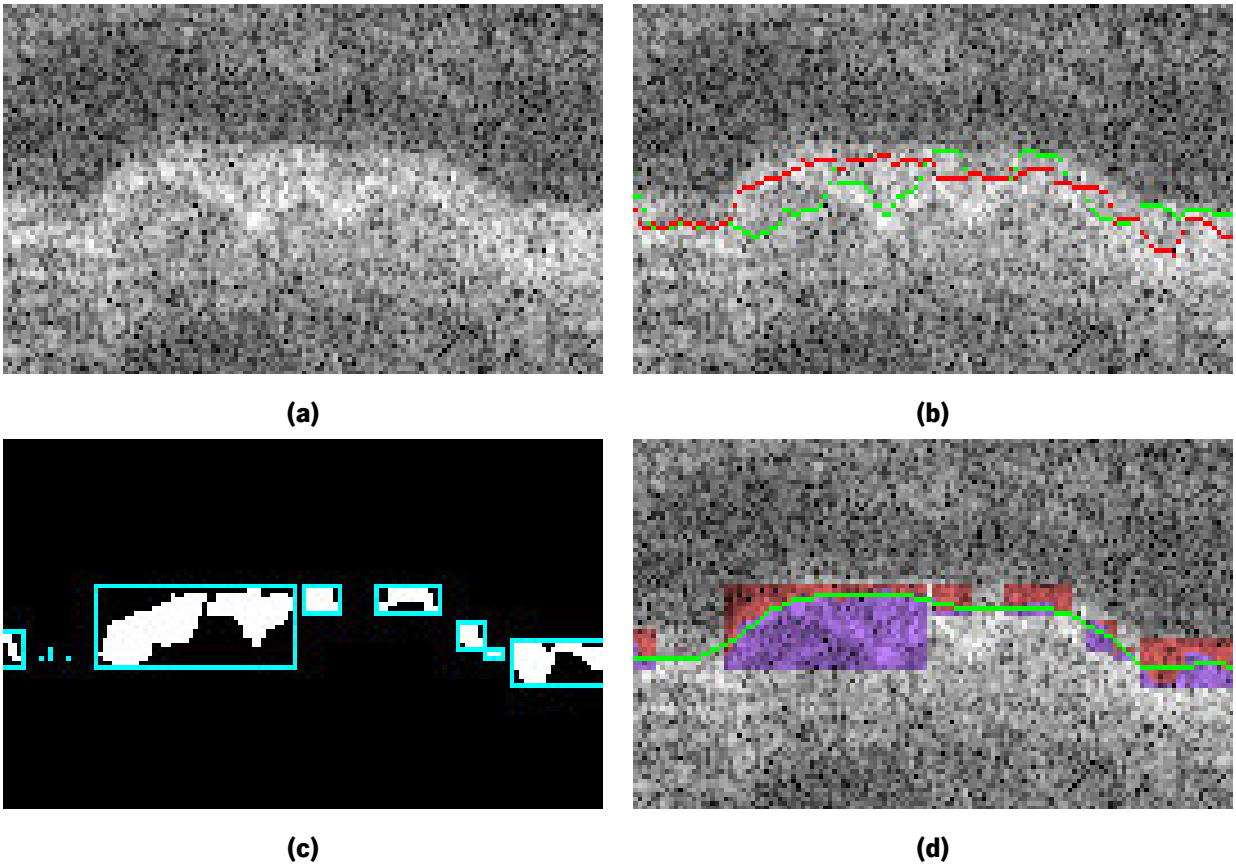


Figure 5.12: Detection and application of SHOPs. The original image is presented in (a). In (b), the IRPEDC from step 6 is presented in green, while the IRPEDC without soft constraints from step 7 is depicted in red. Image (c) refers to the difference between these two segmentations, which produces objects used to define the cliques for each SHOP (rectangles in cyan). The objects consider a connectivity of 8 pixels. Image (d) shows in green the IRPEDC segmentation using SHOPs (step 8) after the post-processing and the favored labelings ($\boldsymbol{L}_{IRPEDC}$) in red and purple representing labels 0 and 1, respectively.

5.1.2.8 Step 8: IRPEDC resegmentation with SHOPs

Rother $et\ al.$ [100] used SHOPs for a problem of texture restoration, where they applied SHOPs in every pixel with the labelings of the most repeated patterns. Consequently, the computation complexity was considerably increased. In this case, the aim is not to recover reoccurring patterns, instead the goal is to apply a few SHOPs to improve the segmentation on some particular locations. This implies that the number of SHOPs applied in each image is low and that the computational complexity is just slightly increased relative to the framework without SHOPs[5].

SHOPs act as a local shape prior that encourages the boundary to follow the set of labelings $\boldsymbol{L}_{IRPEDC}$ (Fig. 5.11 and 5.12). SHOPs are applied in the form of compact representation, which encourages labelings similar to each favored labeling L_{IRPEDC} with slightly higher penalizations (Fig. 5.1). Moreover, SHOPs penalties are

[5]Processing times are presented in appendix A.2

standardized for different clique sizes, as explained in section 5.1.1.4.

Apart from the use of SHOPs, this step is similar to the previous joint segmentation of the IS-OS and the IRPEDC from step 6.

5.1.2.9 Step 9: Post-processing

Manual segmentations of boundaries tend to have a slight offset from the actual boundary [53, 128, 77]. Therefore, these offsets were learned from the manual segmentations of expert 1 of the training set and were used to correct the segmentations when testing. In the end, each boundary is smoothed with a mean filter to resemble the smoothness observed in manual segmentations[6].

5.2 Experimental setup

5.2.1 Database

The proposed algorithm was evaluated in a publicly available database of 20 patients with intermediate AMD [85]. The database is divided into 4 groups of patients according to the type of lesions present and the quality of the images (Table 5.1). The quality of the images differ mainly in terms of spatial resolution[7] and presence of motion artifacts (Fig. 5.20) [85]. These volumes were acquired by 4 SD-OCT imaging systems from Bioptigen, Inc. with a mean pixel size of 3.19 x 6.56 μm in the axial and lateral directions, respectively. For each volume, 11 slices are available with a mean distance between them that varies from 135 to 337.5 μm. The central slice is located on the fovea. The database also includes manual segmentations of the ILM, the IRPEDC and the BM boundaries performed by 2 expert graders. More detailed information may be found in [85]. To evaluate results in drusen regions, a manual detection of drusen regions was performed by an expert grader. The grader was asked to mark the limits, in the x direction, of each drusen region.

Table 5.1: Composition of the database.

	Group 1	**Group 2**	**Group 3**	**Group 4**
Lesions	Drusen	Drusen	Drusen and GA	Drusen and GA
Image quality	Good	Good	Bad	Bad
Number of volumes	5	5	5	5

[6] Examples of manual segmentations in appendix B

[7] The term resolution refers to the ability to differentiate two objects (i.e., spatial resolution) and not to the number of pixels.

5.2.2 Evaluation

The metric selected to evaluate the performance of the algorithms for each individual boundary was the unsigned error (or absolute error). The automatic segmentations were evaluated by considering the manual segmentations of each expert grader and the average of experts' segmentations. The results of each A-scan were considered as a single observation, implying that the mean and standard deviation of these metrics are relative to all A-scans of all slices.

The unsigned error of the proposed method was compared statistically to the results of other methods. This comparison was performed with a Wilcoxon signed-ranked test — a non-parametric statistical test for paired samples [5]. This test evaluates the hypothesis of the results obtained with the proposed method being statistically different from those of another method. The statistical significance was defined by a p-value < 0.05. The test was two-sided and consequently, it was possible to determine if the results of a particular method are statistically higher or lower than those of the proposed method. To perform the statistical analysis, it could also be used the paired t-test [5]. This test assumes that the population follows a normal distribution, which is something that the Wilcoxon signed-ranked does not require [5]. This assumption was the reason why the t-test was not used, since it was empirically observed that the distributions of the unsigned error do not resemble normal distributions. For performing a statistical analysis of unpaired data, it was used the Wilcoxon ranked sum test under the same conditions mentioned before. This test is similar to the Wilcoxon signed-ranked test, but it is used for situations where the samples are not matched (e.g., comparing results between images with good and bad quality).

5.2.3 Parameter definition

The results of the proposed algorithm were evaluated using a 5-fold cross validation, where the training set was composed by 4 folds and the test set by 1 fold. The folds were stratified, implying that each fold contained one OCT volume from each group, resulting in folds with 4 OCT volumes. The training set is used to create the prior dataset, which consists in a collection of height values retrieved from the manual segmentations, as mentioned in subsection 5.1.1.2. With the prior dataset, several parameters of the smoothness and interaction constraints (subsections 5.1.1.2 and 5.1.1.3) can be computed through simple calculations. However, the variables λ_i (Eq. 5.3), α_{ij} (Eq. 5.6) and p (Eq. 5.10) cannot be computed for the context of an individual energy term. These variables are responsible for the scaling of each energy term, meaning that their values are correlated (e.g., a decrease of α_{ij} may be compensated by an increase of λ_i). Therefore, the effect of these variables on the whole energy must be evaluated simultaneously. With that aim, they were learned through a Bayesian optimization algorithm [125][8].

The Bayesian optimization algorithm [125] models the loss function as a Gaussian process. This model provides the best set of parameters for the next iteration according to a particular criterion. In this case, the criterion is the expected reduction in differential entropy [125]. The Bayesian optimization was selected because

[8]The range of values for each hyperparameter is exposed in appendix A.1

it avoids a blind and exhaustive search of parameters, i.e., a grid search.

The Bayesian optimization [125] was performed through a 4-fold cross validation on the training set (3 folds for training and 1 fold for validation). The search limits of the hyperparameters were empirically selected. Nonetheless, it was taken into account that the search space limits should not be too restrictive to avoid excluding the optimal hyperparameters. Furthermore, the maximum number of evaluations for each algorithm step were also defined empirically by observing when the optimization loss function stabilized. The loss function was the mean unsigned error relative to expert 1 of every boundary considered in a particular segmentation[9].

5.2.4 Dufour $et\ al.$ algorithms implementations

The algorithms proposed by Dufour $et\ al.$ [1] for healthy retinas and for retinas with drusen were both implemented in this work. These algorithms were used to assess the importance of developing an algorithm that takes into account the presence of drusen and GA. The algorithm for healthy retinas is presented in Fig. 5.13 and the algorithm for retinas with drusen is delineated in Fig. 5.14. In the third step of the algorithm for healthy retinas, the original implementation of Dufour $et\ al.$ [1] does not segment the IRPEDC, however it was assumed that if it did, it would perform a joint segmentation of the 3 boundaries. The first two steps of both implementations were performed exactly as the proposed algorithm to perform a fair comparison of the IRPEDC and the BM, which are the main focus of this work. The search spaces were defined as in [1] and the confidence level for defining the hard constraint limits was 99% ($\beta = 2.6$) for all steps [1], except the first two that were performed exactly as the proposed method.

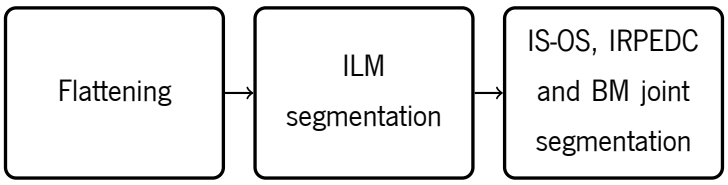


Figure 5.13: Diagram of the algorithm proposed Dufour $et\ al.$ [1] for healthy retinas.

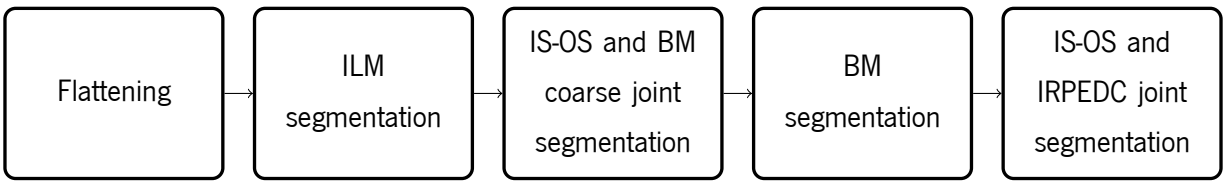


Figure 5.14: Diagram of the algorithm proposed by Dufour $et\ al.$ [1] for retinas with drusen.

[9]The database does not include manual segmentations of the IS-OS. Without the manual segmentation it is not possible to compute the error, thus this boundary was always disregarded from the loss function.

5.3 Results and discussion

In this section, several tests were performed to evaluate the performance of the proposed algorithm. First, the error was computed for the whole database to obtain an overall performance measure. Afterwards, the results were analyzed in more detail for situations such as: drusen and non-drusen regions, volumes with and without GA and for volumes of good and bad image quality.

The evaluations included 4 methods besides the proposed: the proposed method without SHOPs, to evaluate the contribution of SHOPs, Dufour's methods for healthy retinas and retinas with drusen [1], to evaluate the importance of the structure of the algorithm, and Chiu's method [85], to perform a fair comparison with a state of the art method in a common public database [85].

5.3.1 Analysis of overall errors

The boundary errors for the whole database are presented in Table 5.2. The proposed method performs similarly to an expert grader for all boundaries, as can be verified by comparing the error of the proposed method for expert 1 and 2 with the intergrader variability (Table 5.2).

For the ILM, the proposed algorithm performs significantly better than Chiu's method for expert 1 and the average of experts, but worse for expert 2 (Table 5.2). Chiu's method does not include any kind of regularization for segmentation of the ILM. This could explain the better result of this method relative to expert 2. The regularization may hinder the segmentation, since the ILM exhibits considerable height variations, mainly caused by the fovea depression. Moreover, the ILM is a very prominent boundary, implying that it can be segmented just with the information of the image (E_{bound} of Eq. 4.36). The proposed method and the methods of Dufour have the same results for the ILM (Table 5.2), since their first two steps were implemented in the same way, as referred earlier in subsection 5.2.4. With the same segmentation of the ILM, the results of the IRPEDC and the BM can be compared more accurately.

The IRPEDC boundary seems to be the most difficult boundary to segment. This is supported by the largest errors and intergrader variability of all boundaries (Table 5.2). For the IRPEDC, the error of the proposed method is almost always significantly lower than Dufour's and Chiu's methods (Table 5.2). The only exception is for Chiu's method in relation to expert 1, in which the error is statistically significantly higher. Furthermore, the SHOPs improved the segmentation of the IRPEDC of the proposed algorithm in a statistically significant manner (Table 5.2).

As for the BM, the proposed method obtains an error significantly lower than all other methods, except when comparing with Dufour's method for drusen and expert 2 is defined as the reference, in which case the proposed algorithm presents a statistically significant higher error (Table 5.2).

5.3.2 Analysis of errors in drusen and non-drusen regions

Drusen cause morphological changes in the retina, in particular for the IRPEDC and BM boundaries. Empirically, it was observed that those morphological changes hinder the segmentation of both boundaries. The information of Table 5.3 also suggests that the IRPEDC and the BM are significantly more difficult to identify in drusen regions than in non-drusen regions.

Considering drusen regions, the proposed method performs significantly better segmenting the IRPEDC than all other methods (Table 5.3). As for the BM, the results of the proposed method in drusen regions are almost always significantly better than Chiu's method and Dufour's method for healthy retinas. Nonetheless, the errors of the proposed method are significantly higher than those of the other method of Dufour for retinas with drusen (Table 5.3). For non-drusen regions, the proposed method attains the lowest errors in most of the cases (Table

Table 5.2: Overall mean unsigned error ($\pm$ standard deviation) in μm. The lowest mean values for each expert are presented in bold. Statistically significant results greater or lower than those of the proposed method are represented by $\uparrow$ and $\downarrow$, respectively. The absence of these symbols indicates no statistical difference between results. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data).

Method	Grader	ILM	IRPEDC	BM
Chiu $et\ al.$ **[85]**	1	3.39 $\pm$ 2.93 $\uparrow$	**6.32** $\pm$ 6.91 $\downarrow$	4.19 $\pm$ 4.48 $\uparrow$
	2	**3.79** $\pm$ 3.56 $\downarrow$	7.36 $\pm$ 7.41 $\uparrow$	5.35 $\pm$ 4.98 $\uparrow$
	avg	3.05 $\pm$ 2.60 $\uparrow$	5.86 $\pm$ 6.25 $\uparrow$	4.18 $\pm$ 4.12 $\uparrow$
Dufour $et\ al.$ **[1]**	1	3.26 $\pm$ 2.85 $\uparrow$	8.33 $\pm$ 10.04 $\uparrow$	4.96 $\pm$ 6.99 $\uparrow$
	2	3.93 $\pm$ 3.54 $\uparrow$	8.60 $\pm$ 10.22 $\uparrow$	6.02 $\pm$ 7.35 $\uparrow$
	avg	2.94 $\pm$ 2.69 $\uparrow$	7.50 $\pm$ 9.64 $\uparrow$	4.79 $\pm$ 6.91 $\uparrow$
Dufour $et\ al.$ **[1]**	1	3.26 $\pm$ 2.85 $\uparrow$	7.24 $\pm$ 7.92 $\uparrow$	4.37 $\pm$ 5.47 $\uparrow$
	2	3.93 $\pm$ 3.54 $\uparrow$	7.61 $\pm$ 8.06 $\uparrow$	5.57 $\pm$ 6.04 $\downarrow$
for drusen	avg	2.94 $\pm$ 2.69 $\uparrow$	6.43 $\pm$ 7.25 $\uparrow$	4.32 $\pm$ 5.34 $\uparrow$
Proposed without SHOPs	1	3.26 $\pm$ 2.85	7.62 $\pm$ 8.33 $\uparrow$	4.07 $\pm$ 4.22
	2	3.93 $\pm$ 3.54	8.10 $\pm$ 8.40 $\uparrow$	5.29 $\pm$ 4.89
	avg	2.94 $\pm$ 2.69	6.90 $\pm$ 7.69 $\uparrow$	4.00 $\pm$ 4.00
Proposed	1	**3.26** $\pm$ 2.85	6.39 $\pm$ 6.56	**4.07** $\pm$ 4.22
	2	3.93 $\pm$ 3.54	**6.86** $\pm$ 6.68	**5.29** $\pm$ 4.89
	avg	**2.94** $\pm$ 2.69	**5.53** $\pm$ 5.66	**4.00** $\pm$ 4.00
Intergrader variability		4.09 $\pm$ 3.64	7.30 $\pm$ 6.87	5.03 $\pm$ 4.37

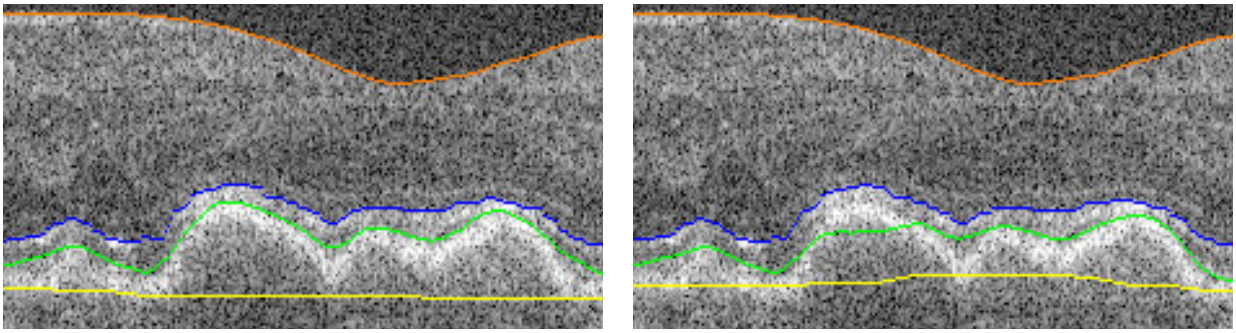
Table 5.3: Mean unsigned error ($\pm$ standard deviation) in μm of drusen regions and non-drusen regions. The lowest mean values for each expert are presented in bold. Statistically significant results greater or lower than those of the proposed method are represented by $\uparrow$ and $\downarrow$, respectively. The absence of these symbols indicates no statistical difference between results. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The results for regions of drusen and non-drusen were also compared statistically with a Wilcoxon ranked sum test (for unpaired data). The region with significantly lower unsigned error was underlined.

		IRPEDC		**BM**	
Method	**Grader**	**Drusen**	**Non-drusen**	**Drusen**	**Non-drusen**
	1	8.89 $\pm$ 9.28 $\uparrow$	**5.74** $\pm$ 6.10 $\downarrow$	5.06 $\pm$ 5.13 $\uparrow$	3.99 $\pm$ 4.30
Chiu $et\ al.$	2	10.28 $\pm$ 9.76 $\uparrow$	6.70 $\pm$ 6.58 $\uparrow$	**5.88** $\pm$ 5.22 $\downarrow$	5.22 $\pm$ 4.92 $\uparrow$
[85]	avg	8.55 $\pm$ 8.76 $\uparrow$	5.25 $\pm$ 5.33 $\uparrow$	4.85 $\pm$ 4.46 $\uparrow$	4.03 $\pm$ 4.02 $\uparrow$
	1	13.18 $\pm$ 15.78 $\uparrow$	7.21 $\pm$ 7.77 $\uparrow$	6.02 $\pm$ 8.43 $\uparrow$	4.71 $\pm$ 6.59 $\uparrow$
Dufour $et\ al.$	2	14.23 $\pm$ 15.92 $\uparrow$	7.32 $\pm$ 7.85 $\uparrow$	7.09 $\pm$ 8.82 $\uparrow$	5.77 $\pm$ 6.94 $\uparrow$
[1]	avg	12.73 $\pm$ 15.62 $\uparrow$	6.30 $\pm$ 7.12 $\uparrow$	5.86 $\pm$ 8.36 $\uparrow$	4.55 $\pm$ 6.51 $\uparrow$
	1	9.12 $\pm$ 10.24 $\uparrow$	6.81 $\pm$ 7.21 $\uparrow$	4.56 $\pm$ 4.76 $\downarrow$	4.32 $\pm$ 5.62 $\uparrow$
Dufour $et\ al.$	2	9.96 $\pm$ 10.14 $\uparrow$	7.07 $\pm$ 7.40 $\uparrow$	5.89 $\pm$ 5.22 $\downarrow$	5.49 $\pm$ 6.21 $\downarrow$
[1] (drusen)	avg	8.44 $\pm$ 9.51 $\uparrow$	5.97 $\pm$ 6.53 $\uparrow$	4.55 $\pm$ 4.30 $\downarrow$	4.27 $\pm$ 5.55 $\uparrow$
Proposed	1	9.69 $\pm$ 10.33 $\uparrow$	7.15 $\pm$ 7.73 $\uparrow$	4.56 $\pm$ 4.65	3.96 $\pm$ 4.11
without	2	10.49 $\pm$ 10.43 $\uparrow$	7.56 $\pm$ 7.76 $\uparrow$	5.88 $\pm$ 5.13	5.15 $\pm$ 4.83
SHOPs	avg	9.00 $\pm$ 9.77 $\uparrow$	6.42 $\pm$ 7.04 $\uparrow$	4.53 $\pm$ 4.19	3.88 $\pm$ 3.95
	1	**7.82** $\pm$ 8.33	6.06 $\pm$ 6.03	**4.56** $\pm$ 4.65	**3.96** $\pm$ 4.11
Proposed	2	**8.80** $\pm$ 8.70	**6.42** $\pm$ 6.04	**5.88** $\pm$ 5.13	**5.15** $\pm$ 4.83
	avg	**7.10** $\pm$ 7.62	**5.17** $\pm$ 5.04	**4.53** $\pm$ 4.19	**3.88** $\pm$ 3.95
Intergrader variability		8.64 $\pm$ 7.65	5.66 $\pm$ 4.70	6.99 $\pm$ 6.65	4.89 $\pm$ 4.28

5.3).

The frequent favorable results of the proposed method in drusen and non-drusen regions are tightly related with designing an algorithm for handling the morphological changes caused by AMD, such as drusen and GA (non-drusen regions), and with using local shape priors in the form of SHOPs.

The importance of having an algorithm specifically developed for AMD is corroborated by the worse results of the method of Dufour's $et\ al.$ for healthy retinas when compared to the method proposed by the same author for retinas with drusen or even to the proposed segmentation method without SHOPs (Table 5.3). The method of Dufour $et\ al.$ for healthy retinas does not take in account the presence of lesions that cause severe morphological changes. Therefore, this method does not perform well for large drusen or extensive GA regions (Fig. 5.15 and 5.19). This is a consequence of violations of some implicit assumptions underlying the MSS problem. Namely,

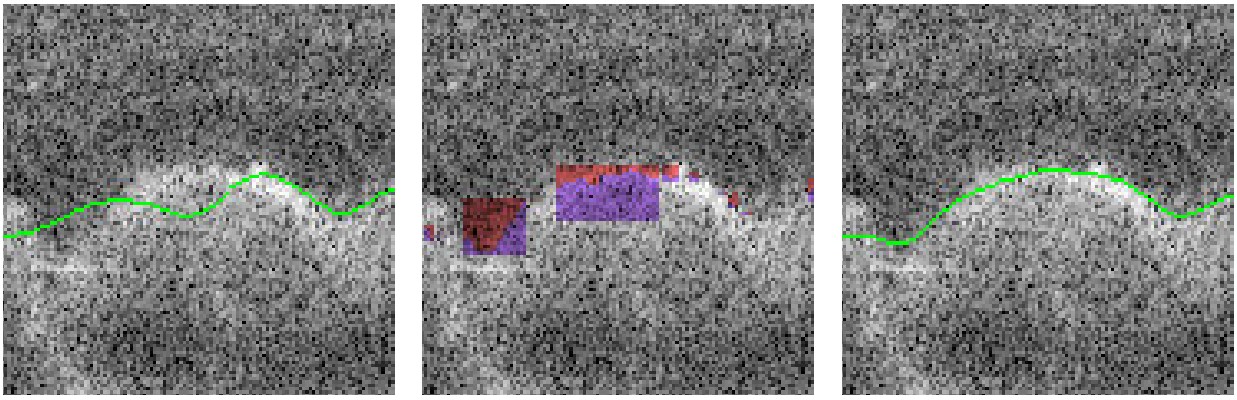


(a) **(b)**

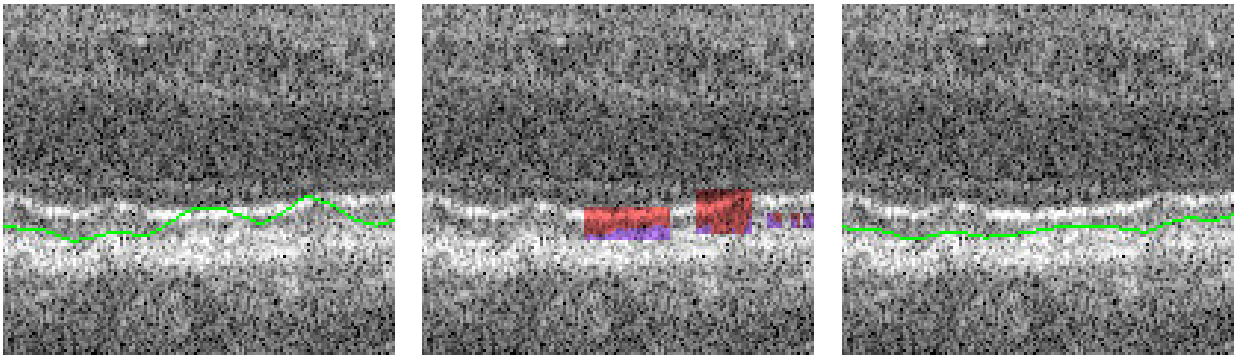
Figure 5.15: Comparison of results of the proposed method without SHOPs (a) against those of Dufour's method (b) in regions of large drusen. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow).

the problem assumes that the boundaries are smooth (smoothness constraints), approximately parallel and the distances between them are roughly constant (interactions constraints). Moreover, the MSS problem also assumes that the number of boundaries is constant throughout the image. These characteristics are generally observed in healthy retinas [54]. In diseased retinas, the morphological changes caused by lesions frequently violate these assumptions [50]. For instance, large drusen disrupt the assumptions of smoothness, near constant distance and parallelism of the IRPEDC and the BM. Thus, the joint segmentation of the IS-OS, the IRPEDC and the BM of the method of Dufour $et\ al.$ for healthy retinas may cause oversmoothing of the IRPEDC and the BM to follow the ORPE (Fig. 5.15 (b)). The proposed algorithm incorporates this knowledge by setting the $\beta = 3.5$ and by segmenting the IRPEDC and the BM independently. The new value of β allows a wider range of height variations for the smoothness constraint, while the independent segmentation avoids the assumptions of near constant distance and parallelism. With this division, the segmentation retains most of the prior knowledge useful for healthy regions and discards the prior information detrimental for lesions, such as the near constant distance between the IRPEDC and the BM (Fig. 5.15 (b)). Other authors have also incorporated the prior knowledge of morphological changes related with AMD by segregating segmentation of the IS-OS, the IRPEDC and the BM into a few steps [1, 85, 50]. For instance, in images with drusen, Dufour et al. [1] divided these three segmentations into three steps (subsection 5.2.4). The authors justify this alteration mainly because the BM segmentation would follow the ORPE in the cases of large drusen [1]. This is partially explained by the lack of contrast of the BM in those regions (Fig. 5.15), as referred by Dufour et al. [1], but also by the violations of the assumptions of parallelism and near constant distant, as mentioned earlier.

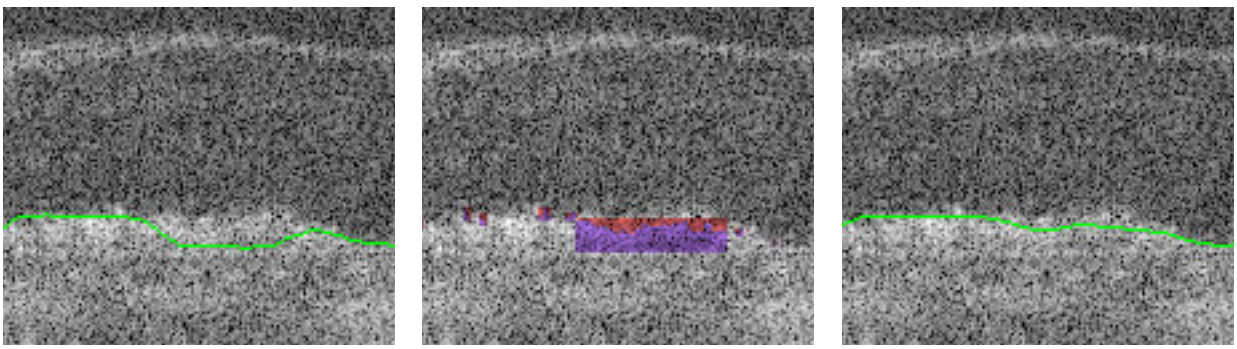
The SHOPs improve the performance for the IRPEDC in both types of regions in a statistically significant manner (Table 5.3). This supports the empirical observations: SHOPs frequently correct the oversmoothing of the IRPEDC boundary in drusen regions (first row of Fig. 5.16). As for non-drusen regions, SHOPs correct the IRPEDC when it is attracted to the IS-OS (second row of Fig. 5.16), because the top limit of the search space used for defining the SHOPs excludes pixels that compose the transition of the IS-OS (subsection 5.1.2.7). Additionally, SHOPs also correct some situations where the IRPEDC is attracted to dark-to-light transitions within the RPE (third



(a) IRPEDC without SHOPs **(b)** Favored labelings of SHOPs **(c)** IRPEDC with SHOPs



(d) IRPEDC without SHOPs **(e)** Favored labelings of SHOPs **(f)** IRPEDC with SHOPs



(g) IRPEDC without SHOPs **(h)** Favored labelings of SHOPs **(i)** IRPEDC with SHOPs

Figure 5.16: Examples of cases where the SHOPs improve the segmentation of the IRPEDC (green). The first row refers to the situation where a drusen is oversmoothed, the second row to the case of the IRPEDC segmentation being attracted to the IS-OS and the third row to the attraction of the IRPEDC to transitions within the RPEDC. The column in the middle shows the favored labelings of SHOPs overlaid in red and purple. The red and purple pixels represent respectively the labels 0 and 1 of the favored labelings.

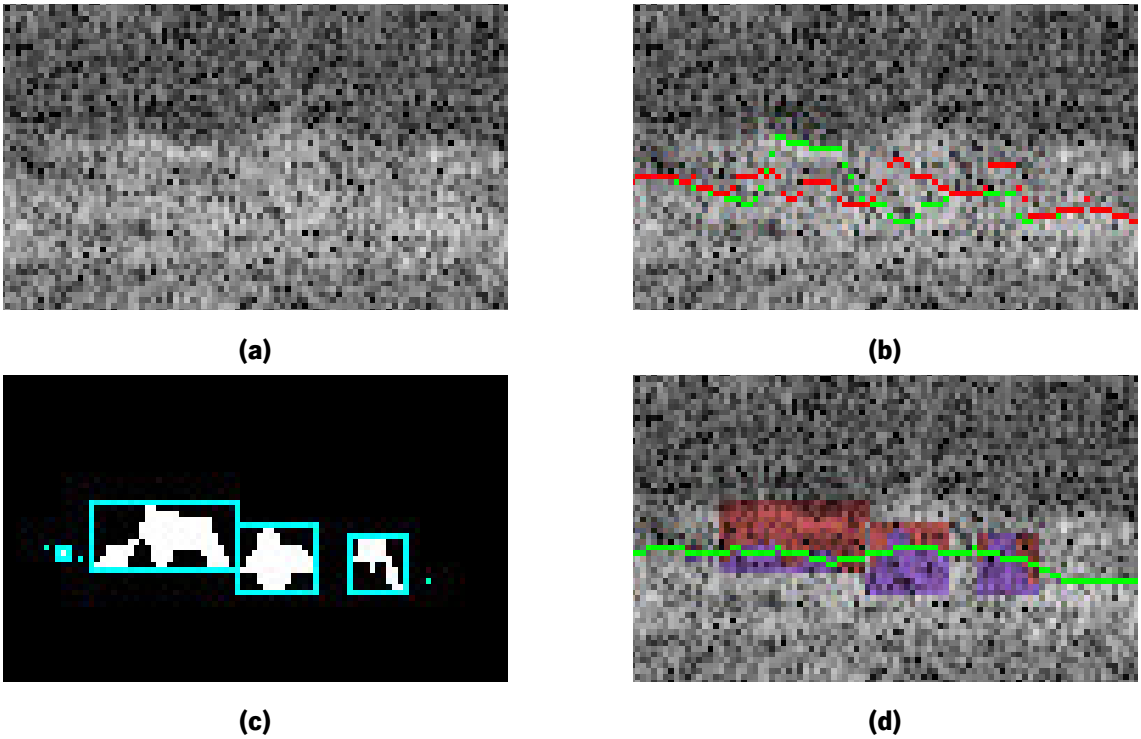


Figure 5.17: Detection and application of SHOPs in a case where SHOPs are incorrectly defined. The original image is presented in (a). In (b), the IRPEDC from step 6 is shown in green, while the IRPEDC without soft constraints from step 7 is presented in red. Image (c) refers to the difference between these two segmentations, which is used to define the cliques for each SHOP (rectangles in cyan). Image (d) shows the IRPEDC segmentation using SHOPs (step 8) after the post-processing in green and the favored labelings ($\boldsymbol{L}_{IRPEDC}$) in red and purple for the labels 0 and 1, respectively.

row of Fig. 5.16). This is mainly related with the lower limit of the search space used for defining the SHOPs (subsection 5.1.2.7), which depends on intensity information rather than the gradient, granting robustness to the definition of SHOPs to situations of low IRPEDC contrast or foci of hyperreflective drusen material. The procedure to compute the SHOPs sometimes provides incorrect labelings, which are mainly related with the less constrained segmentation of the IRPEDC (subsection 5.1.2.7). This segmentation is sensitive to the presence of noise, artifacts and other strong image gradients (Fig. 5.17 (b)). The SHOPs can be interpreted as local soft shape prior, which can encourage the segmentation to follow a predefined shape in a defined neighborhood without imposing it (Fig. 5.17). Hence, incorrectly defined SHOPs are frequently not fully followed to avoid an increase in the cost of the other constraints of the segmentation (Fig. 5.17 (d)). This behavior grants some immunity to the presence of incorrectly defined SHOPs. In sum, SHOPs have stronger modeling capabilities than the other constraints of the framework. Nevertheless, SHOPs are also more complex to apply for two reasons. First, they require specific knowledge about the context in which they are applied. For example, the flexibility added by SHOPs to the segmentation causes an undesired increase of sensitivity to noise or artifacts. To attenuate this issue, SHOPs were integrated into a constrained segmentation. Second, the favored labelings demand knowledge about the problem in order to be defined. If the labelings are not properly defined, they may

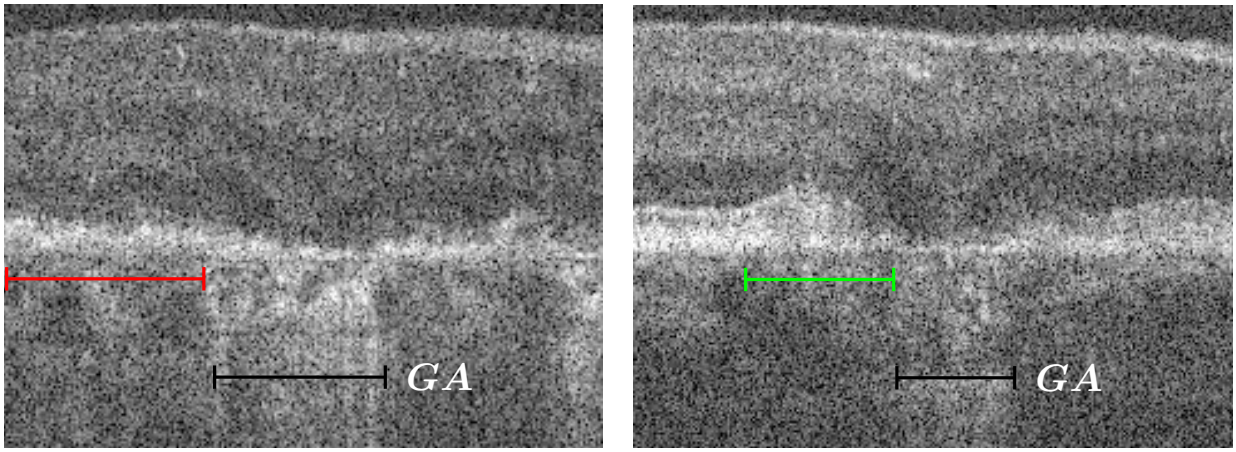


Figure 5.18: Two examples of regions with GA (regions in black). In (a), an example of the irregularity of the IRPEDC near a GA is presented (region in red). In (b), it is shown an example of the thinning of the RPEDC near a GA (region in green).

cause coarse errors that even a constrained segmentation cannot alleviate.

5.3.3 Analysis of the presence of GA

The presence of GA causes several morphological changes to the outer retina. Near the regions of GAs, it was empirically observed that the RPEDC layer is thinner and the IRPEDC boundary is much more irregular (Fig. 5.18). Furthermore, the surrounding tissue is oftentimes unstructured (Fig. 5.18). All these alterations difficult the segmentation of the IRPEDC, which is reflected in a larger intergrader variability and in larger errors for most automatic methods (Table 5.4).

Converse to the previous situation, the segmentation of the BM is enhanced by the alterations related with the GAs. This is corroborated by the improvement of the intergrader variability, as well as of the majority of the results of automatic methods in volumes with GA (Table 5.4). The result improvement is related with the contrast enhancement of the BM caused by the thinning of the RPEDC. The RPEDC is composed by highly reflective tissue, thus its thickness reduction allows more infrared light to penetrate deeper tissues, which in consequence improves visualization of the BM and the choroid (Fig. 5.18).

In general, the proposed method obtained better results for volumes containing GA than the remaining methods (Table 5.4). In fact, the improvements are almost always statistically significant for both the IRPEDC and the BM (Table 5.4). In extensive regions of GA, the method of Dufour for healthy retinas often performs very poorly (Fig. 5.19 (b)). This is expected because this algorithm was developed without considering the presence of any type of lesion [1]. In GA regions, the smoothness, the parallelism and the near constant distance of the IRPEDC and the BM boundaries are roughly maintained (Fig. 5.18). However, the distance between the boundaries is severely reduced, which in conjunction with the enhanced visualization of the choroid may lead to attraction of the segmentation to transitions in the choroid (Fig. 5.19 (b)). Furthermore, the IS-OS is often

missing in GA lesions, violating another implicit assumption of the MSS problem – having of a constant number of boundaries throughout the image. As a consequence, Dufour's method for healthy retinas segments incorrectly the three boundaries — the IRPEDC is segmented in the place of the IS-OS and transitions of the choroid are segmented instead of the IRPEDC and the BM (Fig. 5.19 (b)). To incorporate the information of the lacking IS-OS, the proposed algorithm performs the preliminary (or coarse) segmentation of the IRPEDC (subsection 5.1.2.4) by jointly segmenting just 2 boundaries: the IRPEDC and the BM. This preliminary segmentation is subsequently used for determining the top limit of the search space of the independent segmentation of the BM (step 5, subsection 5.1.2.5). Dufour's method for drusen also performs a preliminary segmentation of the BM using 2 boundaries: the IS-OS and the BM (subsection 5.2.4). However, the search space for independent segmentation of the BM is a band around the preliminary segmentation of the BM [1]. In GA regions, this

Table 5.4: Mean unsigned error ($\pm$ standard deviation) in μm for images with and without GA. The lowest mean values for each expert are presented in bold. Results statistically greater or lower than those of the proposed method are represented by $\uparrow$ and $\downarrow$ respectively. The absence of these symbols indicates no statistical difference between results. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The results for the set of volumes with GA and without GA were also compared statistically with a Wilcoxon ranked sum test (for unpaired data). The set of volumes with significantly lower unsigned error was underlined.

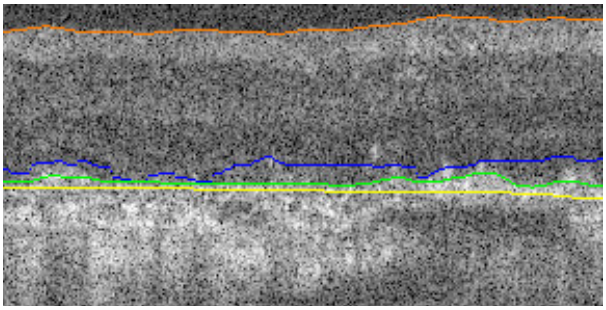
		IRPE		**BM**	
Method	**Grader**	**No GA**	**GA**	**No GA**	**GA**
	1	**5.50** $\pm$ 5.66 $\downarrow$	7.15 $\pm$ 7.89 $\uparrow$	**3.98** $\pm$ 4.07	4.40 $\pm$ 4.86 $\uparrow$
Chiu $et\ al.$	2	6.35 $\pm$ 6.08 $\uparrow$	8.37 $\pm$ 8.42 $\uparrow$	**5.49** $\pm$ 5.13 $\uparrow$	5.20 $\pm$ 4.83 $\uparrow$
[85]	avg	**4.87** $\pm$ 4.98 $\downarrow$	6.85 $\pm$ 7.16 $\uparrow$	**4.11** $\pm$ 4.02 $\uparrow$	4.26 $\pm$ 4.21 $\uparrow$
	1	8.00 $\pm$ 8.89 $\uparrow$	8.65$\pm$ 11.06 $\uparrow$	5.21 $\pm$ 7.73 $\uparrow$	4.70 $\pm$ 6.15 $\uparrow$
Dufour	2	7.71 $\pm$ 8.68 $\uparrow$	9.50 $\pm$ 11.49 $\uparrow$	6.54 $\pm$ 8.29 $\uparrow$	5.50 $\pm$ 6.22 $\uparrow$
[1]	avg	6.90 $\pm$ 8.37 $\uparrow$	8.10 $\pm$ 10.73 $\uparrow$	5.17 $\pm$ 7.82 $\uparrow$	4.42 $\pm$ 5.84 $\uparrow$
	1	6.81 $\pm$ 6.96 $\uparrow$	7.68 $\pm$ 8.75 $\uparrow$	4.17 $\pm$ 4.57 $\uparrow$	4.56 $\pm$ 6.24 $\uparrow$
Dufour $et\ al.$	2	6.78 $\pm$ 6.85 $\uparrow$	8.43 $\pm$ 9.03 $\uparrow$	5.56 $\pm$ 5.43 $\downarrow$	5.57 $\pm$ 6.58
(drusen) [1]	avg	5.79 $\pm$ 6.25 $\uparrow$	7.06 $\pm$ 8.07 $\uparrow$	4.21 $\pm$ 4.51 $\uparrow$	4.43 $\pm$ 6.05 $\uparrow$
Proposed	1	7.45 $\pm$ 7.71 $\uparrow$	7.79 $\pm$ 8.91 $\uparrow$	4.09 $\pm$ 4.32	4.05 $\pm$ 4.13
without	2	7.58 $\pm$ 7.49 $\uparrow$	8.63 $\pm$ 9.19 $\uparrow$	5.52 $\pm$ 5.15	5.05 $\pm$ 4.61
SHOPs	avg	6.56 $\pm$ 7.07 $\uparrow$	7.24 $\pm$ 8.24 $\uparrow$	4.13 $\pm$ 4.22	3.88 $\pm$ 3.77
	1	6.01 $\pm$ 5.76	**6.76** $\pm$ 7.25	4.09 $\pm$ 4.32	**4.05** $\pm$ 4.13
Proposed	2	**6.31** $\pm$ 5.87	**7.42** $\pm$ 7.36	5.52 $\pm$ 5.15	**5.05** $\pm$ 4.61
	avg	5.08 $\pm$ 4.97	**5.98** $\pm$ 6.23	4.13 $\pm$ 4.22	**3.88** $\pm$ 3.77
Intergrader variability		6.90 $\pm$ 6.11	7.70 $\pm$ 7.54	5.12 $\pm$ 4.36	4.94 $\pm$ 4.37

preliminary segmentation of the BM was sometimes attracted to light-to-dark transitions in the choroid and the search space would be far from the BM and would not contain it as consequence (Fig. 5.19 (c) and (d)). This difference may explain the tendency of the proposed method performing better for volumes with GA than the Dufour's method for drusen (Table 5.4).

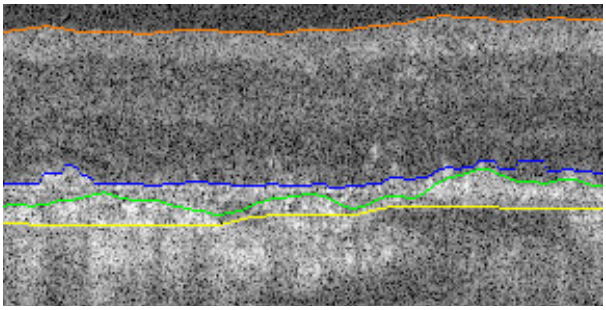
Despite being adapted to images with GA, the proposed method also performs well in images without this kind of lesion (Table 5.4). As can be observed in Table 5.4, the proposed method significantly outperforms most of the other methods in volumes without GA.

5.3.4 Analysis of image quality

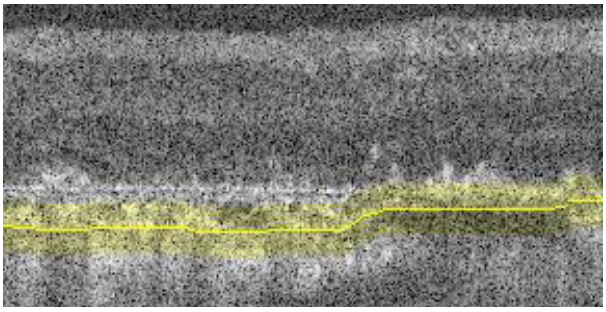
The database is divided into two levels of image quality: good and bad quality (subsection 5.2.1). As mentioned in section 5.2.1, the quality levels differ mainly in terms of spatial resolution and presence of motion artifacts (Fig. 5.20) [85]. The image quality caused two distinct scenarios for the IRPEDC and the BM (Table 5.5). For the IRPEDC, there is no clear trend on how the image quality affects the segmentation. It was not found a statistical significant difference for the intergrader variability (Table 5.5). As for the automatic methods, there are statistical differences, but it was not possible to identify a clear trend (Table 5.5). For the BM, the segmentation performs better in images of good quality. For this level of quality, the variability between manual graders is lower and all



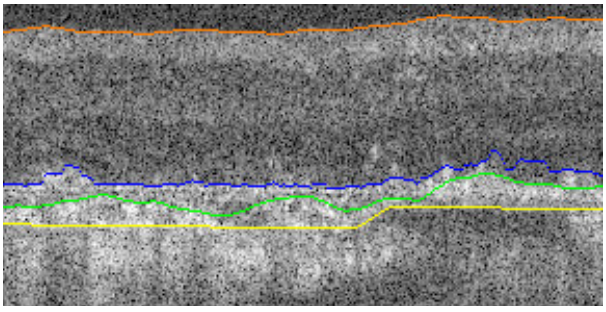
(a) Proposed methods without SHOPs



(b) Dufour's method for healthy retinas



(c) Preliminary BM of Dufour's method for drusen



(d) Dufour's method for drusen

Figure 5.19: Comparison of results in an extensive GA region. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow). In image (c) the yellow boundary is the preliminary (or coarse) segmentation of the BM of Dufour's method for drusen and the yellow band surrounding it corresponds to the search space for the final segmentation of the BM.

Table 5.5: Mean unsigned error ($\pm$ standard deviation) in μm for images with good and bad quality. The lowest mean values for each expert are presented in bold. Results statistically greater or lower than those of the proposed method are represented by $\uparrow$ and $\downarrow$ respectively. The absence of these symbols indicates no statistical difference between results. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The results for the sets of volumes with good and bad image quality were also compared statistically with a Wilcoxon ranked sum test (for unpaired data). The set of volumes with significantly lower results was underlined.

		IRPEDC		**BM**	
Method	**Grader**	**Good quality**	**Bad quality**	**Good quality**	**Bad quality**
	1	**6.19** $\pm$ 6.59 $\downarrow$	6.46 $\pm$ 7.22	4.08 $\pm$ 4.52 $\uparrow$	**4.30** $\pm$ 4.44
Chiu $et\ al.$	2	7.13 $\pm$ 6.92 $\uparrow$	7.60 $\pm$ 7.87 $\uparrow$	5.14 $\pm$ 4.68 $\uparrow$	**5.55** $\pm$ 5.27 $\downarrow$
[85]	avg	5.66 $\pm$ 5.81 $\uparrow$	6.06 $\pm$ 6.65 $\uparrow$	4.10 $\pm$ 4.01 $\uparrow$	**4.27** $\pm$ 4.22 $\uparrow$
	1	8.17 $\pm$ 8.98 $\uparrow$	8.48 $\pm$ 11.00 $\uparrow$	4.38 $\pm$ 5.83 $\uparrow$	5.53 $\pm$ 7.94 $\uparrow$
Dufour	2	8.08 $\pm$ 9.08 $\uparrow$	9.13 $\pm$ 11.22 $\uparrow$	5.33 $\pm$ 6.06 $\uparrow$	6.70 $\pm$ 8.39 $\uparrow$
[1]	avg	7.12 $\pm$ 8.51 $\uparrow$	7.88 $\pm$ 10.64 $\uparrow$	4.20 $\pm$ 5.63 $\uparrow$	5.38 $\pm$ 7.94 $\uparrow$
	1	7.45 $\pm$ 7.77 $\uparrow$	7.03 $\pm$ 8.06 $\uparrow$	4.31 $\pm$ 5.97 $\uparrow$	4.43 $\pm$ 4.93 $\downarrow$
Dufour $et\ al.$	2	7.49 $\pm$ 7.67 $\uparrow$	7.72 $\pm$ 8.43 $\uparrow$	5.35 $\pm$ 6.36 $\downarrow$	5.79 $\pm$ 5.69 $\downarrow$
(drusen) [1]	avg	6.44 $\pm$ 7.03 $\uparrow$	6.41 $\pm$ 7.45 $\uparrow$	4.22 $\pm$ 5.82 $\uparrow$	4.42 $\pm$ 4.80 $\downarrow$
Proposed	1	8.10 $\pm$ 8.28 $\uparrow$	7.14 $\pm$ 8.36 $\uparrow$	3.75 $\pm$ 3.65	4.39 $\pm$ 4.71
without	2	8.25 $\pm$ 8.03 $\uparrow$	7.96 $\pm$ 8.75 $\uparrow$	4.87 $\pm$ 4.34	5.70 $\pm$ 5.36
SHOPs	avg	7.22 $\pm$ 7.53 $\uparrow$	6.58 $\pm$ 7.83 $\uparrow$	3.69 $\pm$ 3.40	4.32 $\pm$ 4.50
	1	6.38 $\pm$ 6.26	**6.39** $\pm$ 6.84	**3.75** $\pm$ 3.65	4.39 $\pm$ 4.71
Proposed	2	**6.51** $\pm$ 5.97	**7.21** $\pm$ 7.31	**4.87** $\pm$ 4.34	5.70 $\pm$ 5.36
	avg	**5.31** $\pm$ 5.13	**5.76** $\pm$ 6.13	**3.69** $\pm$ 3.40	4.32 $\pm$ 4.50
Intergrader variability		7.23 $\pm$ 6.75	7.36 $\pm$ 6.99	4.74 $\pm$ 4.11	5.32 $\pm$ 4.60

automatic methods perform better regardless of the grader used as reference (Table 5.5).

In relation to other methods, the proposed algorithm attains a significantly lower error for the IRPEDC for both levels of image quality, being the only exception the method of Chiu when the grader 1 is selected as reference. As for the BM boundary, the proposed method performs almost always better than the other methods in images of good quality (Table 5.5). However, for bad quality images, the results of the proposed method are not so competitive; it performs better than Dufour's method for healthy retinas, worse than Dufour's method for drusen and for Chiu's method, the performance is very dependent on the grader selected as reference (Table 5.5).

5.3.5 Analysis of signed errors

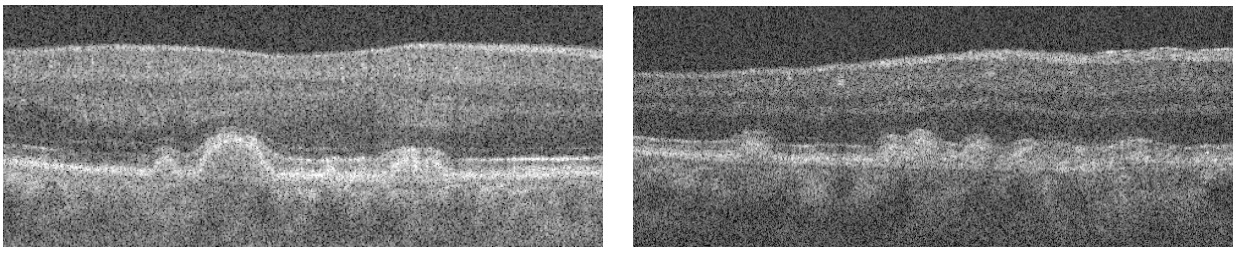
The global results in terms of mean signed error reveal that the boundaries present a slight offset relative to the manual segmentation (Table 5.6). In particular, the IRPEDC exhibits considerable offsets, which is expected because it is also is the most difficult boundary to segment. As mentioned in subsection 5.1.2.9, the expert graders have the tendency to segment the boundary with a small bias, which might explain the offsets observed in Table 5.6. To minimize this issue, the mean offsets were learned from a training set and corrected at test time (all the previous results included the offset correction).

Table 5.6: Mean signed error ($\pm$ standard deviation) in μm without any offset correction.

Method	Grader	ILM	IRPEDC	BM
Chiu $et\ al.$	1	1.87 $\pm$ 4.07	-1.61 $\pm$ 5.69	-1.39 $\pm$ 5.98
	2	1.09 $\pm$ 5.08	-3.89 $\pm$ 9.70	-1.51 $\pm$ 7.15
	avg	1.48 $\pm$ 3.72	-2.75 $\pm$ 8.11	-1.45 $\pm$ 5.69
Proposed	1	-0.66 $\pm$ 4.35	-1.38 $\pm$ 9.16	-2.88 $\pm$ 5.82
	2	-1.44 $\pm$ 5.25	-3.65 $\pm$ 9.34	-3.00 $\pm$ 7.19
	avg	-1.05 $\pm$ 3.98	-2.51 $\pm$ 7.86	-2.94 $\pm$ 5.63
Intergrader variability		-0.78 $\pm$ 5.42	-2.27 $\pm$ 9.76	-0.12 $\pm$ 6.66

5.4 Summary

This chapter started by exposing how the MSF can be adapted to segment retinal boundaries. The external boundary energy term was based on the intensity gradient and, in a few times, on the intensity. The limits of the hard constraints for both the smoothness and interaction energy terms were defined through the height distributions of the prior dataset, being these distributions modeled as Gaussian functions. The soft constraints of the smoothness and interaction energy terms were set as penalizations on the deviation from the mean values



(a) Good quality **(b)** Bad quality

Figure 5.20: Example of images from volumes of good and bad quality.

of height variation and distance between boundaries, respectively. Additionally, the local shape priors were standardized for different sizes.

Afterwards, the proposed algorithm was presented. The algorithm starts by attenuating the speckle noise and by flattening the images to support a correct boundary segmentation in the following steps. The ILM is the first boundary to be segmented. It is an auxiliary boundary, nevertheless, it is important for restricting the search space for the segmentation of the IRPEDC and the BM – the relevant boundaries for this work. Thereafter, the algorithm segments the IRPEDC and the BM in several steps instead of a joint segmentation. After segmenting the IRPEDC and the BM, the algorithm resegments the IRPEDC with local shape priors. At the end, the boundaries are smoothed and offset to approximate the automatic segmentations to those traced by expert graders.

Thereupon, the algorithm was tested in database of 20 patients. The algorithm was compared to methods of Chiu $et\ al.$ [85] and Dufour $et\ al.$ [1] in terms of mean unsigned error. For most cases, the proposed algorithm performs better than the other methods evaluated. The proposed method maintains that trend for regions of drusen, images containing GA and images of bad quality. These results are mainly justified by the structure of the algorithm and by the usage of local shape priors. The algorithm was developed to avoid situations where the implicit assumptions of MSS problem would be violated, e.g., drusen and GA. The local shape priors were applied as SHOPs that were capable of improving the segmentation of the IRPEDC, mostly in regions of drusen oversmoothing and incorrect attraction of the IRPEDC to the IS-OS boundary.

Chapter 6

Detection of drusen

In the previous chapter, it was proposed an algorithm for segmenting boundaries that limit drusen. In this chapter, it will be proposed a method for detecting drusen. The location of drusen and the respective segmentation of the IRPEDC and the BM can be used for quantifying drusen material, which is the main objective of this work (section 1.2). The quantification of drusen is important because it is related with the progression of AMD, as mentioned earlier in section 2.2.1.

The proposed detection method is based on the method proposed by Dufour $et\ al.$ [6]. This method detects drusen by using a RF to classify each column of a B-scan as drusen or non-drusen. It detects the majority of drusen, nevertheless it has some limitations. More particularly, this method cannot differentiate contiguous drusen and does not include local information[1]. To overcome these limitations, the current work proposes a method that uses new features, performs a multi-label classification and uses a hybrid HMM-RF model. The hybrid model combines a RF into a HMM, thus allowing the incorporation of local information into the classification.

The method of Dufour $et\ al.$ [6] and the proposed detection method will be characterized in the following section. In section 6.2, the experimental conditions for evaluating the proposed method are outlined. At last, the results of the proposed method will be presented and discussed in section 6.3. Here, each contribution of the proposed method will be evaluated independently. Furthermore, the proposed method will be compared with other methods present in the literature.

6.1 Methods

Most of the drusen detection methods present in the literature are based on thresholding the distance between the BM and the ORPE or a surrogate boundary[2] [4, 87, 18, 5, 81]. Thresholding may detect the vast majority of drusen, but it is likely to ignore drusen or portions of these lesions with reduced height. Furthermore, thresholding

[1] The features used for the classification include local information, however the classification in itself does not. This means that the label of a sample is not influenced by the labels of the neighboring samples.

[2] A surrogate boundary is necessary because the ORPE does not always corresponds to a light-to-dark transition, as mentioned earlier in section 3.2.

may also consider other morphological changes as drusen, such as regions with large RPEDC thickness. Dufour $et\ al.$ [6] proposed an alternative approach that detects drusen through classification. The use of classification simplifies the use of several sources of information to detect drusen. This approach uses several textural features from a neighborhood above the BM, contrasting with the thresholding approach, which relies solely on the distance among the boundaries limiting drusen. Still, the approach proposed by Dufour $et\ al.$ [6] is not capable of differentiating contiguous drusen. With individual detection of drusen, it could be explored a new range of metrics that go beyond the total drusen area or volume [129], such as drusen size, used to categorize the stages of AMD [27] or the evolution of the amount of drusen material per lesion [129]. Furthermore, the method of Dufour $et\ al.$ [6] does not include local information. In this problem, the local information can be useful because the spatial distribution of classes is smooth, i.e., a column of a particular class is very likely to be surrounded by columns of the same class. To address these limitations of the method of Dufour $et\ al.$ [6], the proposed detection method introduces new features based on the distance among boundaries limiting drusen and performs classification with 3 and 4 classes. Additionally, the proposed method uses a RF to model the emission distribution of a HMM, thus creating a hybrid HMM-RF model that classifies columns using local information.

6.1.1 Dufour $et\ al.$ method

The method of Dufour $et\ al.$ [6] locates drusen using a RF to classify columns as drusen or non-drusen. The features are the outputs of a subset of the Leung and Malik filter bank (LMFB), which was originally developed for representing textures [86]. The subset of LMFB is composed by filters of the first and second derivatives of Gaussians for 6 orientations and 2 scales (24 filters), Gaussian filters for 4 scales and Laplacian of Gaussian filters for 8 scales[3] (Fig. 6.1) [86, 130]. The outputs of the LMFB are collected for 12 predefined sampling positions of the neighborhood of each column [6]. The sampling points are located at the BM and at fixed distances above the BM (Fig. 6.2). In healthy tissue, the positions above the BM are meant to correspond approximately to the IRPEDC, the IS-OS and tissue above these boundaries (Fig. 6.2 (a)). However, in drusen regions, this correspondence is no longer valid, which allows the discrimination of drusen and non-drusen columns (Fig. 6.2 (b)).

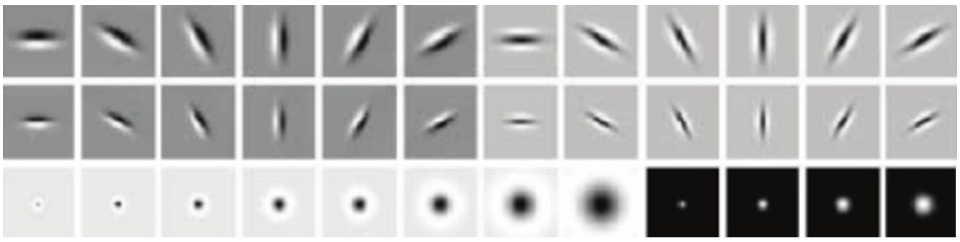
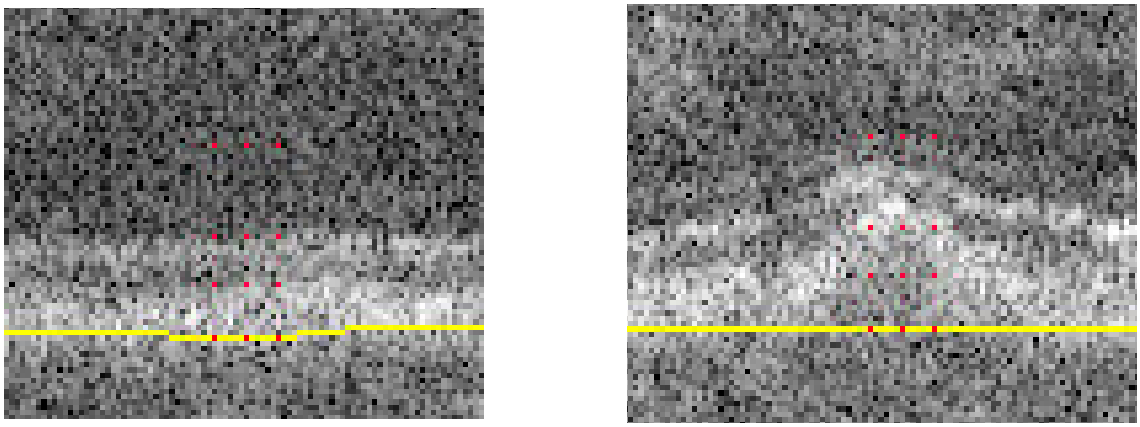


Figure 6.1: Subset of the Leung and Malik filter bank used in method of Dufour $et\ al.$ [6]. Kernels of the filters of the first and second derivatives of Gaussians are presented in the first two rows, while the kernels of the Laplacian of Gaussian and Gaussian filters are in the bottom row (image adapted from [130]).

[3] In this work, it was used the implementation of LMFB available in [130]



(a) Healthy region

(b) Drusen region

Figure 6.2: Sampling positions for the LMFB (red pixels). In (a) the heights of the sampling positions correspond approximately to the BM (yellow pixels), IRPEDC, IS-OS and tissues above the IS-OS. In (b), this correspondence is no longer valid due to the accumulation of drusen material below the RPE.

6.1.2 Proposed method

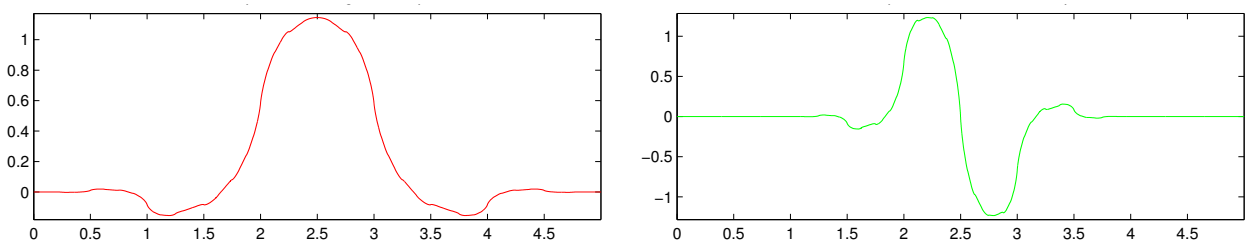
The proposed method differs from the method of Dufour $et\ al.$ [6] in the features, in the number of classes used for the classification and in the classifier itself. The proposed method uses the outputs of the LMFB as features, just as the Dufour $et\ al.$ [6], and it adds some new features, which will be detailed in subsection 6.1.2.1. The outputs of the LMFB were computed in image denoised with a median filter. The denoising was necessary because the images used in this work have considerably more speckle noise than those used by Dufour $et\ al.$ [6]. The median filter was selected because it preserves edges and it is robust to multiplicative noises [126, 127], just as mentioned before in subsection 5.1.2.1. Regarding the number of classes, the proposed method performs the classification with more than two classes. The additional classes will characterized in subsection 6.1.2.2. As for the classifier, the proposed method uses a hybrid HMM-RF model, which consists in the integration of a RF into a HMM. The operations necessary to perform this integration will presented in subsection 6.1.2.3.

6.1.2.1 New features

One of the new proposed features is the distance between the IRPEDC and the BM boundaries, hereinafter designated as $dist_{IRPEDC\_BM}$ (Fig. 6.4 (b)). This feature is based on the fact that drusen material accumulates between the RPE and BM. The value of $dist_{IRPEDC\_BM}$ is not solely dependent on the quantity of drusen material, it also includes the thickness of the RPE layer. Nevertheless, it has been reported that elevations of the IRPEDC relative to the BM are relevant and also correlate with the risk of AMD progressing to an advanced stage [81].

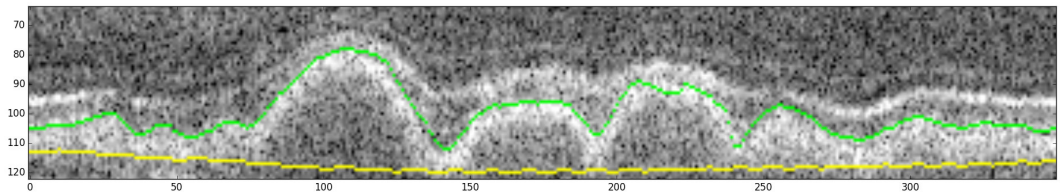
The remaining new features are the wavelet coefficients (WCs) of the $dist_{IRPEDC\_BM}$ from scale 1 to 64 (Fig. 6.4 (c)). The wavelet transform is used for this problem because it can capture spectral characteristics of

a signal over a restricted space [131]. The fact that the information is restricted in space is very important for segmentations tasks. Moreover, the multiple scales can capture spectral information of the elevations caused by drusen of different sizes and shapes. To compute the WC, it was used the continuous wavelet transform. The scaling and wavelet functions used were the $biortogonal1.3$[4] (Fig. 6.3). These functions were empirically selected due to their similarity with drusen shape and good preliminary results.

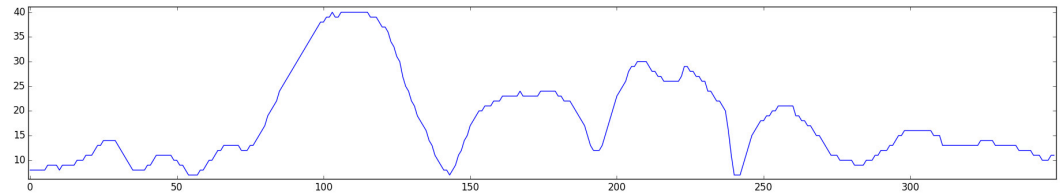


(a) Scaling function **(b)** Wavelet function

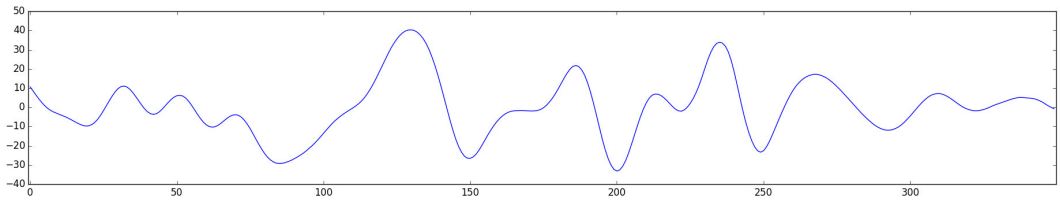
Figure 6.3: $Biortoghonal1.3$ scaling and wavelet functions.



(a) Image with drusen. Depicted boundaries are IRPEDC (green) and BM (yellow)



(b) Distance between IRPEDC and BM ($dist_{IRPEDC\_BM}$)



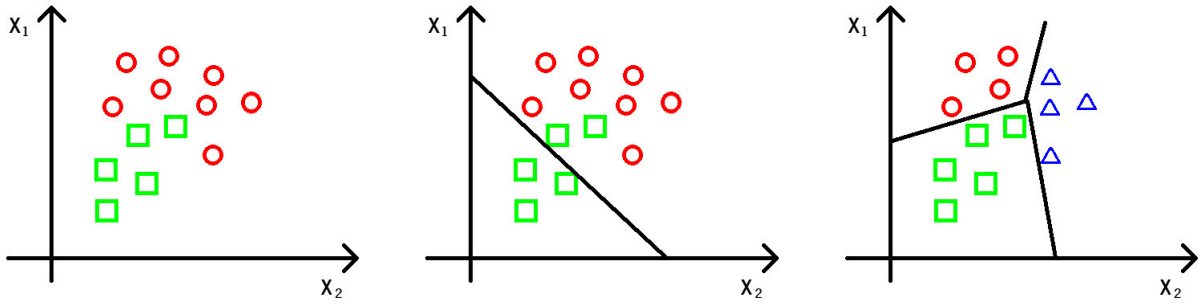
(c) Wavelet transformation of the $dist_{IRPEDC\_BM}$ using the $biortogonal1.3$ wavelet for scale 19

Figure 6.4: Examples of new features for a image containing drusen. The scale 19 is presented here because it is one of the most relevant scales, according to the feature importances determined with a RF.

[4]The wavelet and scaling functions belong to the biorthogonal family and have order 1 for reconstruction and order 3 for decomposition.

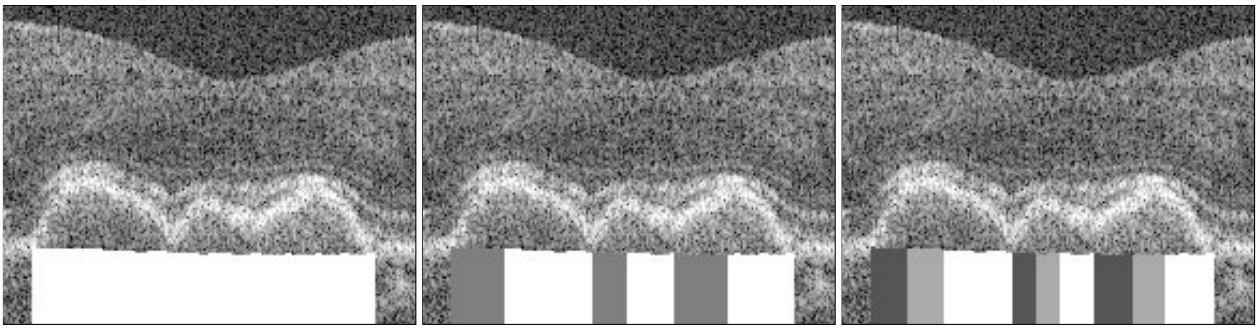
6.1.2.2 Multi-label classification

Besides adding new features, the classification was also performed with 3 and 4 classes (Fig. 6.6). In these situations, the original drusen class is separated into 2 or 3 subclasses (Fig. 6.6). This class was divided because the morphology of drusen varies throughout their extent. Consequently, the features produce distinct outputs to different regions of drusen. For instance, in Fig. 6.4 (c), the wavelet coefficients have different signs for the left and right sides of drusen. This results in a drusen class composed by samples with very heterogeneous feature values. Thus, the classifier needs to create a loose decision boundary to include most of the samples of the drusen class (Fig. 6.5 (b)). A loose decision boundary can induce mistakes because it can easily include samples from the non-drusen class (Fig. 6.5 (b)). However, with a multi-label classification, the drusen class can be subdivided into smaller and more homogeneous subclasses (Fig. 6.5 (c)). This way, the classifier can define tighter decision boundary for the subclasses, which may improve the discrimination of the original classes of drusen and non-drusen (Fig. 6.5 (c)).



(a) Data **(b)** Binary classification **(c)** Multi-label classification

Figure 6.5: An example of improvement with multi-label classification. The data is divided into 2 classes, depicted by green squares and red circles, and it is characterized by features X_1 and X_2. The black lines represent the decision boundaries of the binary and the multi-label classifications. In (c), the class of red circles was subdivided into 2 subclasses, where the blue triangles represent a new subclass. The multi-label classification discriminates fully the 2 original classes of (a).



(a) 2 classes **(b)** 3 classes **(c)** 4 classes

Figure 6.6: Example of using 2, 3 and 4 classes for drusen detection. The labels of drusen subclasses of each column are encoded in the intensity values of the pixels below the BM.

In the classification with 3 classes, drusen are subdivided into 2 subclasses whose frontier is the point of maximum $dist_{IRPEDC\_BM}$ (Fig. 6.6 (b)). To reduce the influence of abnormal segmentations, the point of maximum $dist_{IRPEDC\_BM}$ is defined as the mean location of the 25% largest values of $dist_{IRPEDC\_BM}$. Thus, the 2 drusen subclasses represent drusen regions that essentially exhibit either positive or negative derivatives of $dist_{IRPEDC\_BM}$. For the case of 4 classes, drusen are subdivided into 3 subclasses. The new drusen subclass corresponds to a region surrounding the maximum $dist_{IRPEDC\_BM}$ that occupies about 25% of each drusen (Fig. 6.6 (c)). This subclass is normally characterized by low values of derivatives of $dist_{IRPEDC\_BM}$ and high values of $dist_{IRPEDC\_BM}$, where the latter values differentiate this subclass from the non-drusen class. The other two drusen subclasses will be the regions to the left and right of this new subclass. The subclass in the left contains mainly positive values of the derivatives of the $dist_{IRPEDC\_BM}$, while the subclass in the right is composed primarily by negative values of the derivatives of the $dist_{IRPEDC\_BM}$.

With these definitions of the drusen subclasses, the multi-label classification also presents the advantage of detecting drusen individually when they are contiguous to other drusen, like in the situations of drusen clusters (Fig. 6.6).

6.1.2.3 Hybrid HMM-RF model

Conventional classifiers, such as RF, do not model the relations between neighboring samples, i.e., the class of a particular sample is not influenced by the classes of the neighboring samples. This problem is often addressed by introducing context aware features. Still, the class label of a sample could diverge abruptly from those of the neighboring samples. In this case, this means that a non-drusen label can occur in the middle of a drusen, which causes errors in determining the limiting columns of these lesions. To incorporate local information into the classification, the RF was integrated into a HMM, forming a hybrid HMM-RF model.

Before creating a hybrid HMM-RF model, the HMM must be adapted to the classification task by defining the features vectors as observations and the hidden states as classes. Furthermore, the 3 distributions that compose a HMM[5] must also be computed for context of drusen detection. The initial state[6] and transition[7] distributions are learned through the maximum likelihood estimates. Since the classification is performed in a supervised setting, this implies that the distributions are computed through a normalized counting of the state frequencies from the training set [132]. The emission distribution[8] is modeled indirectly by a RF. The RF is integrated in the HMM as proposed by Renals $et\ al.$ [115] (section 4.2.3.1), i.e., by converting the $a\ posteriori$ distribution (RF output) into a likelihood distribution.

In this thesis, three different transition distributions were computed (Fig. 6.7). These distributions differ in number of states (or classes) – 2, 3 or 4 states were considered. The drusen states were defined to be spatially sequential, therefore the HMM forbids transitions that never occur in the training set. For instance, state D_1

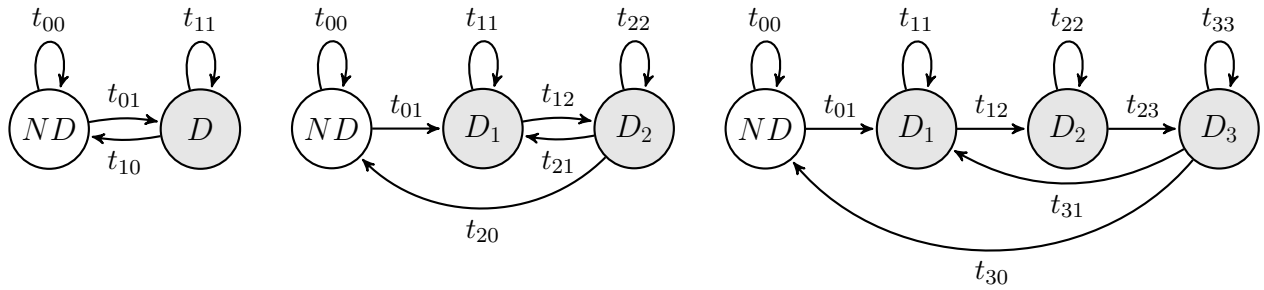
[5]A HMM is composed by 3 distributions: initial state, transition and emission distributions, as mentioned in section 4.2.2.

[6]The initial state distribution is relative to the hidden state of the first sample.

[7]The transition distribution represents the relation between hidden states of neighboring samples.

[8]The emission distribution refers to the relation between individual observations and hidden states.

cannot be followed by state ND (Fig. 6.7 (b) and (c)) because transitioning from the middle of a drusen to a healthy region is not physically possible.



(a) 2 states **(b)** 3 states **(c)** 4 states

Figure 6.7: Graphical representations of the transition distributions of the HMM-RF model for different number of states. Each node corresponds to a different state. ND corresponds to a non-drusen state, D refers a drusen state and D_1, D_2, D_3 are states representing parts of drusen. The value t_{ij} of arc (i, j) is the probability of transitions between nodes i and j.

6.2 Experimental setup

6.2.1 Database

The dataset used for evaluating the detection of drusen is the same used for evaluating the segmentation of the IRPEDC and BM boundaries, which was previously described in section 5.2.1. Table 6.1 presents the number of drusen in the database according to their size. Three size ranges were defined: small, intermediate and large. These were based on the ranges of AREDS for grading the stages of evolution of AMD [27, 29] (section 2.2.1). Drusen in the small range have a width less than 63 μm (about 10 pixels); intermediate drusen exhibit a width equal or superior to 63 μm and inferior to 125 μm, i.e., the range is between 10 and 18 pixels; large drusen present widths of more than 125 μm (more than 18 pixels).

Table 6.1: Number of drusen presented in the database. Drusen are categorized according to their size: small drusen have widths with less than 10 pixels, intermediate drusen present widths between 10 and 18 pixels and large drusen possess widths larger than 18 pixels.

Drusen size	**Small**	**Intermediate**	**Large**	**Total**
Quantity	12	149	646	807

6.2.2 Evaluation

The results of drusen detection will be reported in terms of column classification, number of detected drusen and the drusen area measurements.

The results of the classification of columns were evaluated through the following metrics:

$$Sensitivity = \frac{TP}{TP + FN}, \qquad Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{6.1}$$

$$Specificity = \frac{TN}{TN + FP}, \qquad Dice = \frac{2TP}{2TP + FP + FN}, \tag{6.2}$$

where TP, TN, FP, FN, refer to the number of true positives, true negatives, false positives and false negatives, respectively. Additionally, the area under the ROC curve (AUC) was also determined. These metrics assume that the classification output is binary. Consequently, the results of the multi-label classification had to be computed by reverting the drusen subclasses into a single drusen class.

To compute the number of detected drusen, the results of the classification were converted into locations of the leftmost and the rightmost columns of drusen regions. The rightmost columns were determined by searching for a negative gradient values on the classification output. Each leftmost column was defined in relation to the rightmost column as the farthest column to the left still possessing a drusen label. Besides the number of detected drusen, it will also reported the number of correctly detected drusen. A drusen was assumed to be correctly detected if both the left and right columns were at a distance of 3 pixels or less from the respective manual markings.

The areas of drusen are delimited vertically by the segmentations of the IRPEDC and the BM and horizontally by the locations of drusen. The measurements of those areas were evaluated through the following metrics: sensitivity, dice coefficient, positive predictive value (PPV) and absolute area difference (AAD). The last two metrics have the following definition:

$$PPV = \frac{TP}{TP + FP}, \qquad AAD = |Area_1 - Area_2|, \tag{6.3}$$

where $Area_1$ and $Area_2$ correspond to the drusen areas of each B-scan computed through two different methods. All the metrics to evaluate drusen areas were computed for each B-scan, implying that the mean of each metric is the mean for all the B-scans. Moreover, some of those metrics were undefined when drusen were not present because they included a division by 0. In those cases, it was assumed that the metric value was 0, so the mean and standard deviations could be computed. Metrics such as the specificity and accuracy were not included in this case because these would tend to produce values very close to 100% and with very low variability. This occurs because the metrics of drusen area measurements need to be computed for the whole image, which leads to a very large discrepancy between the amount of true negatives and the remaining variables. The results of drusen area measurements were analyzed statistically. The statistical analysis was performed through a Wilcoxon signed-ranked test, just as in chapter 5. All the conditions of the analysis were the same as mentioned in subsection 5.2.2, i.e., the test is paired and statistical significance was defined by a p-value < 0.05.

6.2.3 Parameter definition

The RF was defined and trained by taking into account the requirements for estimating the $a\ posteriori$ distribution (section 4.2.3.1), which could then be used for integrating the RF into a HMM. One of those requirements consists in having a training set representative of the whole data (section 4.2.3.1), thus the data was divided into 5 stratified folds. Each fold contained one OCT volume from each group, resulting in folds with 4 OCT volumes. The training sets were formed by 3 folds, while the validation and test sets were composed by 1 fold each. This reduces the variability between sets and ensures that the RF is not overajusted to the characteristics of the training and validation sets. Another requirement imposes that the RF must be complex enough to model the $a\ posteriori$ distribution (section 4.2.3.1). Therefore, each RF is composed by 100 trees grown using Gini Impurity[9]. RFs with more trees had very slim gains of performance in the preliminary tests. The number of features considered for each split and the depth of trees were learned through a cross validation scheme that minimizes the cross entropy[10]. This learning procedure was performed through a Bayesian optimization algorithm [125], just as in the segmentation of retinal boundaries (section 5.2.3). The limits for each hyperparameter and the maximum number of iterations were defined empirically. Just as before (section 5.2.3), the hyperparameter limits were selected to avoid excluding the global optimum and the number of iterations was defined by observing when loss function stabilized.

The method of Dufour $et\ al.$ [6] uses the importance of features retrieved from the RF to select the most important 150 features[11], from the 432 initial features. This procedure aims to reduce computational complexity of the method, while retaining a similar performance. The method proposed in the current work also includes feature selection. The number of features is reduced in the same proportion as in the method of Dufour $et\ al.$ [6][12] to perform a fair comparison.

6.3 Results and discussion

All the results presented in the next sections use manual segmentations of the IRPEDC and the BM from expert 1 for computing the features, except if mentioned otherwise. With manual segmentations, the effect of the alterations of the proposed detection method can be evaluated without the interference of the inherent errors of automatic segmentations.

In this section, the results of drusen detection will be exposed and discussed. First, the effect of using new features and multi-label classification will be evaluated. Thereupon, it will be evaluated the effect of incorporating

[9]The Gini Impurity is used to evaluated the quality of the splits when learning each decision tree.

[10]Minimizing a metric such as the cross entropy or the mean squared error is another requirement for a classifier to estimate the $a\ posteriori$ distribution section 4.2.3.1.

[11]These 150 features are considered the be the most important according to the mean Gini Impurity decrease of the whole forest, as mentioned at the end of section 4.2.1.

[12]When using all the new features the total number of features is 497. In this case, the feature selection reduces the amount of features from 497 to 173, thus keeping approximately the same ratio between initial and final amounts of features.

local information to the classification by combining the RF with a HMM. Afterwards, it will be assessed the effect of using the segmentations of expert 2 and the automatic method proposed in chapter 5. At last, the proposed detection method is combined with the proposed segmentation method (chapter 5) and it is evaluated against two other methods present in the literature.

6.3.1 Evaluation of new features and multi-label classification

The results of using multi-label classification and adding new features are presented in Tables 6.2, 6.3, 6.4 and 6.5. The results of using just the LMFB and two classes were also included for comparison. This method is very close to the method of Dufour $et\ al.$ [6], the only difference is that the RF was trained by optimizing the cross entropy instead of the AUC[13].

The addition of any combination of new features improves the overall performance of the classifier, as indicated by the accuracy, dice and AUC in Table 6.2. The improvement in classification of columns also causes an enhancement of mean values of practically every metric regarding drusen area measurements (Table 6.5).

The $dist_{IRPEDC\_BM}$ is a feature that plays an important role on correctly detecting the left and right columns of drusen (e.g., leftmost drusen of the left column in Fig. 6.8), independently of the number of classes. This is corroborated by the improvements of the dice coefficient (Table 6.2) and by the increase in the number of drusen that were correctly detected when adding this feature (Table 6.4). Furthermore, it was observed that this feature can help the classification in regions exhibiting an appearance similar to drusen, but with reduced RPEDC thickness – possible false positives – and regions of drusen with abnormal appearance in which the $dist_{IRPEDC\_BM}$ suggest accumulation of drusen material – possible false negatives (right column of Fig. 6.8).

The WCs of $dist_{IRPEDC\_BM}$ improved the individual detection of drusen when they occur in clusters or very close to each other (Fig. 6.9). This is corroborated by the increase of the number of detected drusen for 3 and 4 classes when adding these features to the LMFB or even to the combination of LMFB and $dist_{IRPEDC\_BM}$ (Table 6.3). Furthermore, the WCs generally enhance the column classification and the measurements of drusen area, particularly for 3 and 4 classes (Tables 6.2 and 6.5). For 3 or 4 classes, the WCs contain spectral information that correlates well with the drusen subclasses, which in turn improves the capability of the RF to differentiate the classes of drusen.

The $dist_{IRPEDC\_BM}$ and the respective WCs can be combined to the LMFB to incorporate the benefits of each type of feature (Fig. 6.9). Consequently, the classifier generally performs better with these features than when the $dist_{IRPEDC\_BM}$ or the respective WCs are added separately (Tables 6.2 and 6.4). In terms of drusen area measurements, using all the features leads to results that are almost always significantly better or at least indifferent than the other combinations of features (Table 6.5).

Besides the combination of features, the number of classes also has a considerable impact on the results. As mentioned in section 6.1.2.2, dividing the drusen class into subclasses allows the classification to detect drusen individually when these occur in clusters (Fig. 6.6). This is reflected in the larger number of detected

[13]The differences in results of training the RF with cross entropy and AUC are reported in appendix C.

Table 6.2: Results of column classification using different combinations of features and number of classes. The feature $dist_{IRPEDC\_BM}$ is represented by D and the respective wavelet coefficients by WC. The values in bold refer to the best results of each row.

	Features	LMFB	LMFB+D	LMFB+WC	LMFB+D+WC
2 classes	**Sensitivity**	0.6258	0.6353	0.6355	**0.6377**
	Specificity	0.9603	0.9607	0.9608	**0.9608**
	Accuracy	0.8949	0.8971	0.8971	**0.8976**
	AUC	0.9149	0.9218	0.9248	**0.9301**
	Dice	0.6996	0.7071	0.7073	**0.7090**
3 classes	**Sensitivity**	0.5976	0.6124	0.6201	**0.6346**
	Specificity	**0.9697**	0.9688	0.9667	0.9644
	Accuracy	0.8969	0.8991	0.8989	**0.8999**
	AUC	0.9163	0.9236	0.9254	**0.9310**
	Dice	0.6940	0.7036	0.7058	**0.7126**
4 classes	**Sensitivity**	0.5737	0.5885	0.6035	**0.6137**
	Specificity	**0.9752**	0.9747	0.9717	0.9695
	Accuracy	0.8966	0.8992	0.8992	**0.8999**
	AUC	0.9155	0.9226	0.9239	**0.9308**
	Dice	0.6847	0.6954	0.7007	**0.7057**

Table 6.3: Results relative to the number of detected drusen using different combinations of features and number of classes. The feature $dist_{IRPEDC\_BM}$ is represented by D and the respective wavelet coefficients by WC.

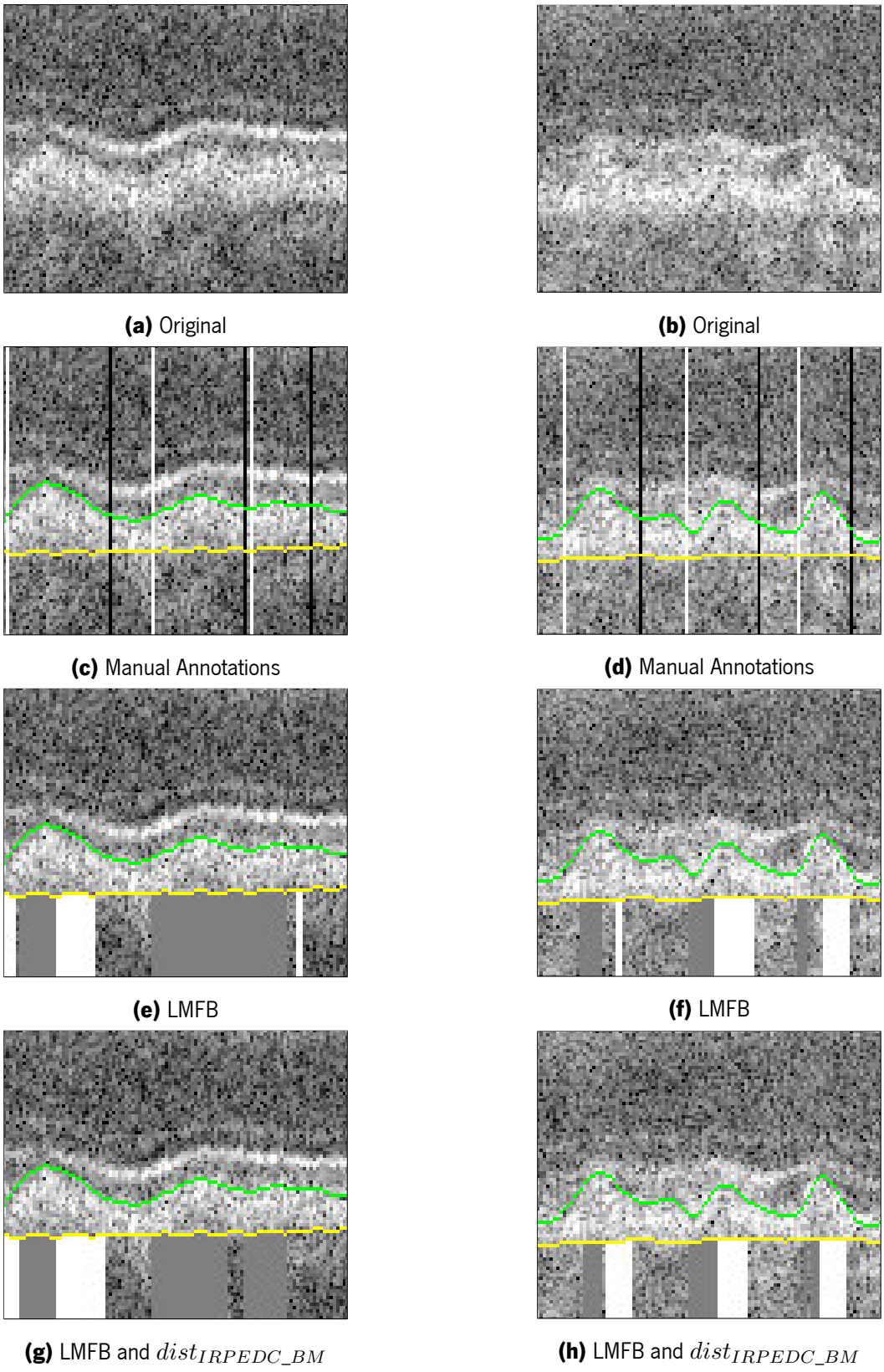
	Features	LMFB	LMFB+D	LMFB+WC	LMFB+D+WC	Expert grader
2 classes	**Small**	490	474	562	411	12
	Intermediate	109	119	117	112	149
	Large	381	379	381	396	646
	Total	980	972	1060	919	807
3 classes	**Small**	604	582	630	610	12
	Intermediate	136	141	168	152	149
	Large	396	402	389	423	646
	Total	1136	1125	1187	1185	807
4 classes	**Small**	561	600	722	645	12
	Intermediate	133	130	139	130	149
	Large	380	385	380	407	646
	Total	1074	1115	1241	1182	807

Table 6.4: Results relative to the number of correctly detected drusen using different combinations of features and number of classes. The feature $dist_{IRPEDC\_BM}$ is represented by D and the respective wavelet coefficients by WC. The values in bold refer to the best results of each row.

	Features	LMFB	LMFB+D	LMFB+WC	LMFB+D+WC
2 classes	**Small**	**1**	**1**	0	0
	Intermediate	21	**27**	19	19
	Large	160	166	155	**175**
	Total	182	**194**	174	**194**
3 classes	**Small**	1	**2**	0	1
	Intermediate	19	19	15	**20**
	Large	194	194	195	**221**
	Total	214	215	210	**242**
4 classes	**Small**	**1**	0	0	**1**
	Intermediate	17	20	14	**21**
	Large	174	181	191	**211**
	Total	192	201	205	**233**

Table 6.5: Results of drusen area measurements using different combinations of features and number of classes The feature $dist_{IRPEDC\_BM}$ is represented by D and the respective wavelet coefficients by WC. The AADs are presented in μm^2. Statistically significant results greater or lower than those of combining all features (LMFB+WC+D) are represented by $\uparrow$ and $\downarrow$, respectively. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The values in bold refer to the best mean results of each row.

	Features	LMFB	LMFB+D	LMFB+WC	LMFB+D+WC
2 classes	**Sensitivity**	0.4972 $\pm$ 0.3781 $\downarrow$	0.5080 $\pm$ 0.3757	0.4977 $\pm$ 0.3737	**0.5099** $\pm$ 0.3790
	PPV	0.6132 $\pm$ 0.4283	0.6241 $\pm$ 0.4238	**0.6250** $\pm$ 0.4223	0.6110 $\pm$ 0.4288
	Dice	0.5149 $\pm$ 0.3766 $\downarrow$	0.5247 $\pm$ 0.3723	0.5194 $\pm$ 0.3707	**0.5259** $\pm$ 0.3762
	AAD	13145.5 $\pm$ 14170.5	**12871.2** $\pm$ 13803.9	12877.0 $\pm$ 13782.6	12899.3 $\pm$ 13891.5
3 classes	**Sensitivity**	0.4666 $\pm$ 0.3709 $\downarrow$	0.4860 $\pm$ 0.3720 $\downarrow$	0.4824 $\pm$ 0.3709 $\downarrow$	**0.5034** $\pm$ 0.3735
	PPV	0.6158 $\pm$ 0.4326 $\uparrow$	0.6229 $\pm$ 0.4308 $\uparrow$	0.6173 $\pm$ 0.4266	**0.6273** $\pm$ 0.4198
	Dice	0.4992 $\pm$ 0.3738 $\downarrow$	0.5158 $\pm$ 0.3747 $\downarrow$	0.5111 $\pm$ 0.3734 $\downarrow$	**0.5254** $\pm$ 0.3723
	AAD	13002.1 $\pm$ 13433.3 $\uparrow$	12494.2 $\pm$ 13051.0 $\uparrow$	12480.3 $\pm$ 12866.6 $\uparrow$	**12376.4** $\pm$ 13028.8
4 classes	**Sensitivity**	0.4456 $\pm$ 0.3686 $\downarrow$	0.4660 $\pm$ 0.3697 $\downarrow$	0.4733 $\pm$ 0.3641 $\downarrow$	**0.4891** $\pm$ 0.3728
	PPV	0.6041 $\pm$ 0.4416	0.6179 $\pm$ 0.4359 $\uparrow$	**0.6335** $\pm$ 0.4237	0.62120 $\pm$ 0.4291
	Dice	0.4858 $\pm$ 0.3779 $\downarrow$	0.5042 $\pm$ 0.3771 $\downarrow$	0.5093 $\pm$ 0.3684 $\downarrow$	**0.5174** $\pm$ 0.3751
	AAD	13195.9 $\pm$ 13481.7 $\uparrow$	12629.9 $\pm$ 13022.1	**12389.2** $\pm$ 12769.9	12585.2 $\pm$ 13002.5



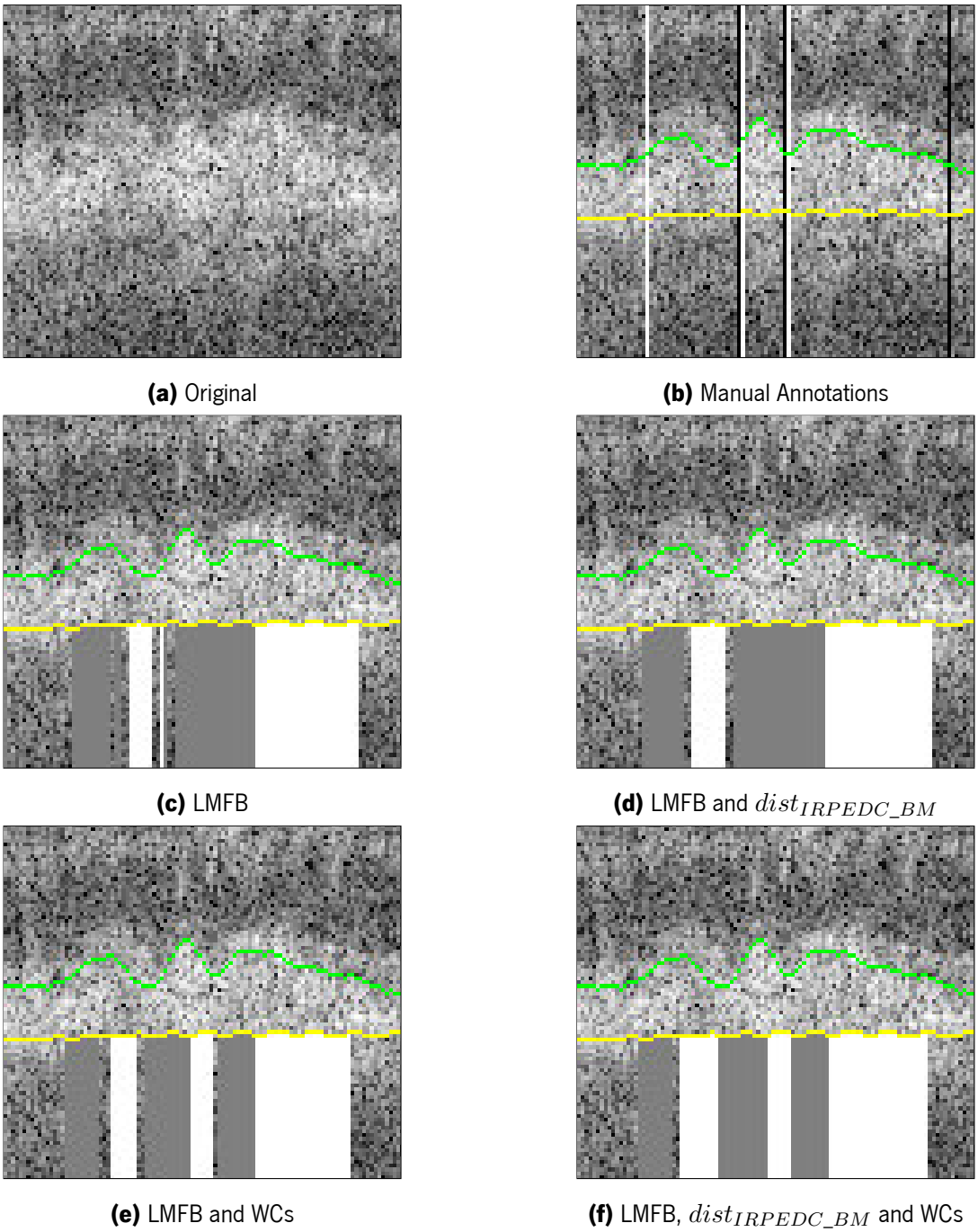
(a) Original **(b)** Original

(c) Manual Annotations **(d)** Manual Annotations

(e) LMFB **(f)** LMFB

(g) LMFB and $dist_{IRPEDC\_BM}$ **(h)** LMFB and $dist_{IRPEDC\_BM}$

Figure 6.8: Example of drusen detection with 3 classes using the LMFB with and without the $dist_{IRPEDC\_BM}$. The manual annotations of the left and right limits of drusen are represented by white and black columns, respectively. The columns marked in white and gray below the BM refer to the different drusen subclasses detected.



(a) Original

(b) Manual Annotations

(c) LMFB

(d) LMFB and $dist_{IRPEDC\_BM}$

(e) LMFB and WCs

(f) LMFB, $dist_{IRPEDC\_BM}$ and WCs

Figure 6.9: Example of drusen detection with 3 classes using different combinations of features. The manual annotations of the left and right limits of drusen are represented by white and black columns, respectively. The columns marked in white and gray below the BM refer to the different drusen subclasses detected.

drusen, correctly or not, when considering more than 2 classes (Table 6.3). Furthermore, the division of the drusen class can improve the classification. For instance, the classification with 3 classes and using all features attains the best results in most metrics (Table 6.2 and 6.5). These results suggest that the data can be divided more coherently by considering more than 2 classes.

6.3.2 Evaluation of the hybrid HMM-RF model

As referred earlier in section 6.1.2.3, the RF classifier can be integrated into a hybrid HMM-RF model to incorporate local information into the classification. In this work, the integration was performed as proposed by Renals $et\ al.$ [115]. This method assumes that the classifier is capable of modeling the $a\ posteriori$ distribution[14] ($p(c|\boldsymbol{v})$), which can then be converted into the likelihood ($p(\boldsymbol{v}|c)$) through the Bayes theorem (Eq. 4.40). The conversion to likelihood was not computed exactly because the evidence term of the Bayes theorem is rather difficult to compute or model. Nonetheless, the likelihood can still be estimated through 3 different ways according to the assumptions made about the statistics of data (more details in section 4.2.3.1):

1. Using directly the $a\ posteriori$ distribution as a estimate of the likelihood,

   $p(\boldsymbol{v}|c) \simeq p(c|\boldsymbol{v})$;
2. Using the $a\ posteriori$ distribution divided by the prior distribution ($p(c)$) as the estimate of the likelihood,

   $p(\boldsymbol{v}|c) \simeq p(c|\boldsymbol{v}) \backslash p(c)$;
3. Using the $a\ posteriori$ distribution computed from balanced data to estimate of the likelihood,

   $p(\boldsymbol{v}|c) \simeq p(c|\boldsymbol{v})_{balanced}$.

The evaluation of the hybrid model was performed for these 3 ways of estimating the likelihood. To simplify this analysis, the features used were the LMFB, the $dist_{IRPEDC\_BM}$ and the respective WCs, since the results of the previous section suggest that this is the best combination of features.

The approach 1 presumes that the $a\ posteriori$ distribution[15] is a good approximation of the likelihood. This assumes that the ratio between the prior and the evidence ($p(v)$) distributions is approximately constant and, therefore, it does not affect the most likely sequence of labels computed by the Viterbi algorithm, as referred earlier in section 4.2.3.1. The $a\ posteriori$ distribution generally attributes higher probabilities to the class of non-drusen, since this class occurs more frequently than the class of drusen[16]. Therefore, the likelihood and subsequently the HMM-RF will be more sensitive to the class of non-drusen than the class of drusen. This is reflected in the higher specificity and lower sensitivity when compared to the other HMM-RF models (Table 6.6). The approaches 2 and 3 try to balance the likelihood probabilities for both classes, thus making these approaches

[14] The is a reasonable assumptions given that the classifier is complex enough, minimizes the cross entropy and the training data is representative of the whole problem (more details in section 4.2.3.1).

[15] A scaled version of the $a\ posteriori$ distribution could also be used.

[16] The problem of drusen detection is naturally imbalanced. The class of non-drusen is roughly 4 times more frequent than the class of drusen.

Table 6.6: Results of column classification for the hybrid HMM-RF model. Three different approaches of integrating the RF into a HMM are presented. The results of the RF classifier were also included for comparison. The values in bold refer to the best results of each row.

		HMM-RF			RF
	Approach	**1) Direct**	**2) Divide prior**	**3) Balanced**	–
2 classes	**Sensitivity**	0.6043	**0.8301**	0.7849	0.6377
	Specificity	**0.9665**	0.8699	0.9027	0.9608
	Accuracy	0.8957	0.8621	0.8797	**0.8976**
	Dice	0.6938	0.7019	**0.7185**	0.7090
3 classes	**Sensitivity**	0.5778	**0.8827**	0.8562	0.6346
	Specificity	**0.9772**	0.8294	0.8531	0.9644
	Accuracy	0.8991	0.8398	0.8537	**0.8999**
	Dice	0.6914	0.6830	0.6959	**0.7126**
4 classes	**Sensitivity**	0.5486	**0.8901**	0.8343	0.6137
	Specificity	**0.9815**	0.8164	0.8723	0.9695
	Accuracy	0.8968	0.8309	0.8647	**0.8999**
	Dice	0.6753	0.6730	**0.7072**	0.7057

Table 6.7: Results of detection of drusen for the hybrid HMM-RF model. Three different approaches of integrating the RF into a HMM are presented. The results of the RF classifier were also included for comparison. The values in bold refer to the best results of each row.

		HMM-RF			RF	Expert grader
	Approach	**1) Direct**	**2) Divide prior**	**3) Balanced**	–	–
2 classes	**Small**	0	0	0	411	12
	Intermediate	28	34	32	112	149
	Large	350	515	463	396	646
	Total	378	549	495	919	807
3 classes	**Small**	0	1	0	610	12
	Intermediate	16	62	31	152	149
	Large	436	995	936	423	646
	Total	452	1058	967	1185	807
4 classes	**Small**	0	2	0	645	12
	Intermediate	13	59	23	130	149
	Large	394	1006	867	407	646
	Total	407	1067	890	1182	807

Table 6.8: Results of correct detection of drusen for the hybrid HMM-RF model. Three different approaches of integrating the RF into a HMM are presented. The results of the RF classifier were also included for comparison. The values in bold refer to the best results of each row.

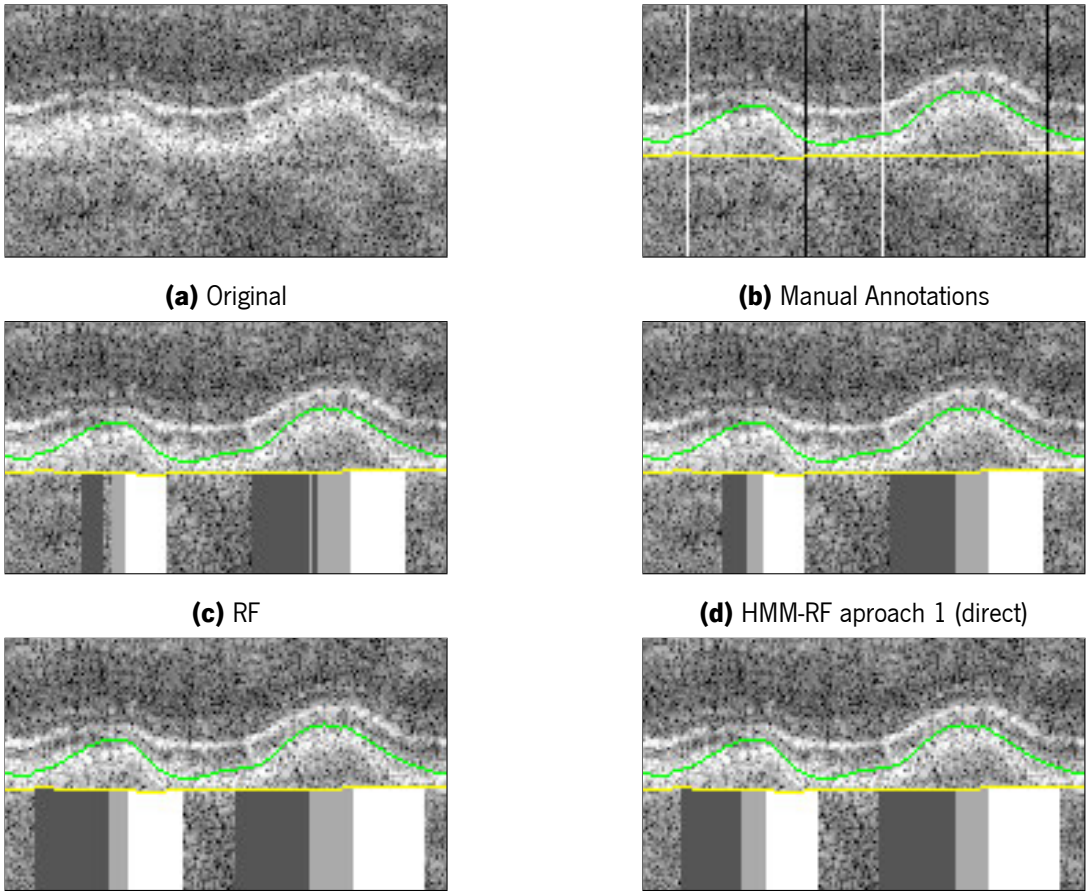
		HMM-RF			RF
	Approach	**1) Direct**	**2) Divide prior**	**3) Balanced**	–
2 classes	**Small**	0	**1**	0	0
	Intermediate	15	**21**	20	19
	Large	132	117	128	**175**
	Total	147	139	148	**194**
3 classes	**Small**	0	**1**	**1**	**1**
	Intermediate	17	21	**25**	20
	Large	230	218	**251**	221
	Total	247	240	**277**	242
4 classes	**Small**	0	**1**	0	**1**
	Intermediate	11	22	**24**	21
	Large	217	217	**253**	211
	Total	228	240	**277**	233

Table 6.9: Results of drusen area measurements for the hybrid HMM-RF model. Three different approaches of integrating the RF into a HMM are presented. The results of the RF classifier were also included for comparison. The AADs are presented in μm^2. Statistically significant results greater or lower than those of the RF classifier are represented by $\uparrow$ and $\downarrow$, respectively. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The values in bold refer to the best results of each row.

		HMM-RF			RF
	Approach	**1) Direct**	**2) Divide prior**	**3) Balanced**	–
2 classes	**Sensitivity**	0.4686 ± 0.3904 ↓	**0.6641** ± 0.4041 ↑	0.6160 ± 0.4093 ↑	0.5099 ± 0.3790
	PPV	0.5724 ± 0.4504	0.5415 ± 0.3767 ↓	0.5597 ± 0.3984 ↓	**0.6110** ± 0.4288
	Dice	0.4920 ± 0.3921 ↓	**0.5677** ± 0.3691 ↑	0.5563 ± 0.3809 ↑	0.5259 ± 0.3762
	ADD	13362.7 ± 14138.2 ↑	16775.4 ± 23245.8 ↑	14287.8 ± 19952.8	**12899.3** ± 13891.5
3 classes	**Sensitivity**	0.4416 ± 0.3820 ↓	**0.7090** ± 0.4013 ↑	0.6886 ± 0.4008 ↑	0.5034 ± 0.3735
	PPV	0.5671 ± 0.4573	0.5135 ± 0.3490 ↓	0.5228 ± 0.3597 ↓	**0.6273** ± 0.4198
	Dice	0.4775 ± 0.3931 ↓	**0.5668** ± 0.3557 ↑	0.5663 ± 0.3588 ↑	0.5254 ± 0.3723
	ADD	13334.9 ± 13294.6 ↑	20330.8 ± 24466.6 ↑	18053.5 ± 22815.7 ↑	**12376.4** ± 13028.8
4 classes	**Sensitivity**	0.4207 ± 0.3779 ↓	**0.7231** ± 0.3932 ↑	0.6752 ± 0.3922 ↑	0.4891 ± 0.3728
	PPV	0.5577 ± 0.4646	0.5073 ± 0.3439 ↓	0.5524 ± 0.3684 ↓	**0.6212** ± 0.4291
	Dice	0.4612 ± 0.3939 ↓	0.5674 ± 0.3496 ↑	**0.5765** ± 0.3579 ↑	0.5174 ± 0.3751
	ADD	14259.8 ± 14204.7 ↑	21378.0 ± 25583.8 ↑	16267.9 ± 22450.7	**125853.2** ± 13002.5

more sensitive to the drusen class and less sensitive to the non-drusen class than the approach 1 (Table 6.6). This behavior is also verified in the drusen area measurements (Table 6.9). The main difference between these approaches is that the approach 2 assumes that the classes have no influence in the evidence distribution and the that the prior distribution from the training set is approximately equal to the prior distribution of the test set, while the approach 3 circumvents these assumptions by learning the $a\ posteriori$ distribution (i.e., training the RF) in a balanced training set, which means that the prior distribution is constant and the evidence distribution can be reasonably assumed to be also constant, since there is no class imbalanced.

The integration of the RF leads to classification results that are generally poorer, as demonstrated by the lower accuracies in Table 6.6. Nonetheless, the incorporation of local information using the hybrid model can improve the detection of drusen in several aspects that justify its use. First, the similarity between the column classification and the manual markings can be improved, as verified by the dice coefficients of approach 3 for 2 and 4 classes (Table 6.6). Second, using the HMM-RF model reduces considerably the number of detected drusen with small and intermediate dimensions (Table 6.7), which is mainly attributable to the merging of drusen



(a) Original **(b)** Manual Annotations

(c) RF **(d)** HMM-RF aproach 1 (direct)

(e) HMM-RF aproach 2 (divide prior) **(f)** HMM-RF approach 3 (balanced)

Figure 6.10: Example of drusen detection with RF and different approaches of HMM-RF. The HMM-RF used 4 classes and all the features. The manual annotations of the left and right limits of drusen are represented by white and black columns, respectively. The columns marked in white and gray below the BM refer to the different drusen subclasses detected.

that were partially detected (Fig. 6.10). Moreover, the number of correctly detected drusen can also be improved, as is the case of employing approach 3 and using more than 2 classes (Table 6.8). Third, the similarity of the delineations of the drusen area computed with the HMM-RF and the manual delineations is significantly improved when using approaches 2 and 3, as indicated by the dice coefficients of Table 6.9. Forth, the HMM-RF forbids physically impossible transitions, such as transitioning from the left class of drusen to the non-drusen class (e.g., left drusen of Fig. 6.10). Furthermore, this incorporation of local information also smooths the outputs of HMM-RF, leading to results to results that are visually closer to the manual markings than those of the RF (Fig. 6.10).

6.3.3 Effect of using automatic segmentations of the IRPEDC and the BM

So far, the features used in the RF were computed using only the manual segmentations of expert grader 1. In this section, it will be analyzed the effect of using other segmentations for that purpose. The use of an automatic segmentation allows evaluating the robustness of the detection method to the presence of segmentation errors. The segmentation method proposed in chapter 5 is used here as the automatic segmentation. Additionally, the segmentation from expert grader 2 is also included to demonstrate how the drusen detection results may vary according to the expert that segments the boundaries.

The evaluation was conducted with the HMM-RF model that utilized 4 classes and the approach 3 for computing the likelihood. Moreover, features employed were the LMFB, the $dist_{IRPEDC\_BM}$ and the respective WCs, like in the previous section. From the analysis of the previous section, it is very difficult to determine the model with the best performance, therefore this model was selected because it holds a reasonable compromise between metrics of different types[17] (Tables 6.6, 6.8 and 6.9).

Employing the automatic segmentation leads to a deterioration of the results in relation to expert 1 (Table 6.10, 6.11 and 6.12). This is expected, as the automatic segmentations always contains some errors (Fig. 6.11). Nonetheless, the results are still comparable to those experts graders. Moreover, the results of automatic segmentation are very similar to those of expert 2 for drusen detection (Table 6.11) and drusen area measurements (Table 6.12). In sum, the results indicate that the drusen detection method exhibits some robustness to segmentation variability, even though the presence of errors can lead to a slight deterioration in performance.

6.3.4 Comparison with methods present in the literature

In this section, the proposed drusen detection method is evaluated against two methods present in the literature. The method of Farsiu $et\ al.$ [133] uses the segmentations of the IRPEDC and the BM of the method Chiu $et\ al.$ [85][18] and detects drusen by thresholding the distance between these boundaries. In [133], the threshold was

[17] This model is also assumed to perform well with other segmentations. This assumption is made because otherwise it would be necessary to perform extensive studies similar to those of sections 6.3.1 and 6.3.2 for each segmentation.

[18] This method was also included in the evaluation of the proposed method of segmentation of drusen limiting boundaries.

Table 6.10: Results of column classification for the hybrid HMM-RF model using different segmentations to compute the features. The values in bold refer to the best results of each row.

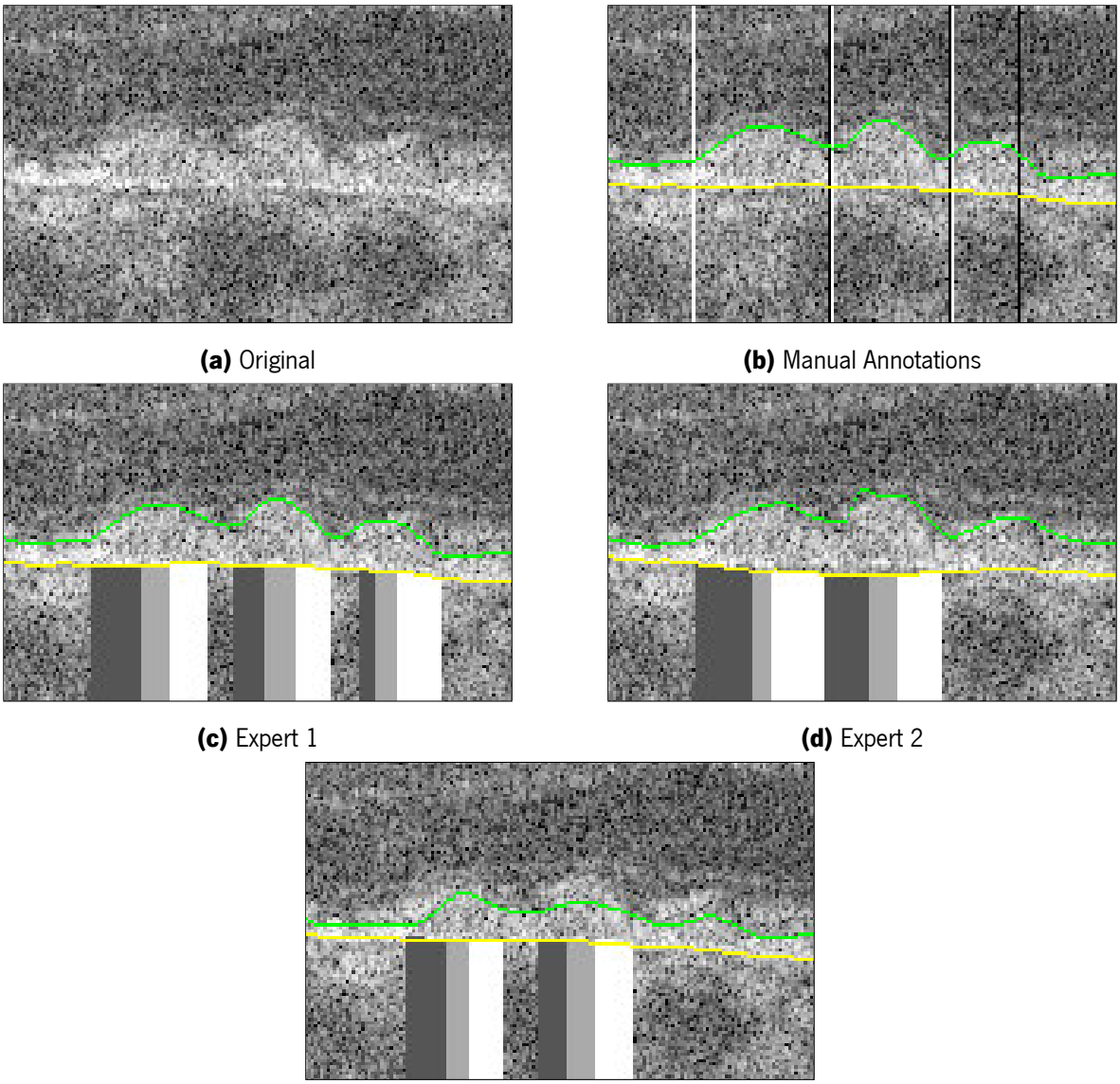
Segmentation	**Expert 1**	**Expert 2**	**Proposed**
Sensitivity	**0.8343**	0.8108	0.8033
Specificity	0.8723	**0.8850**	0.8754
Accuracy	0.8649	**0.8705**	0.8613
Dice	0.7072	**0.7101**	0.6938

Table 6.11: Results of drusen detection for the hybrid HMM-RF model using different segmentations to compute the features. The values in bold refer to the best results of each row.

	Segmentation	**Expert 1**	**Expert 2**	**Proposed**
Detected drusen	**Small**	0	0	0
	Intermediate	23	13	11
	Large	867	734	754
	Total	890	747	765
Correctly detected drusen	**Small**	**0**	**0**	**0**
	Intermediate	**24**	9	8
	Large	**253**	197	196
	Total	**277**	206	204

Table 6.12: Results of drusen area measurements for the hybrid HMM-RF model using different segmentations to compute the features. The AADs are presented in μm^2. Statistically significant results greater or lower than those that use the segmentation of expert 1 are represented by $\uparrow$ and $\downarrow$, respectively. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The values in bold refer to the best results of each row.

Segmentation	**Expert 1**	**Expert 2**	**Proposed**
Sensitivity	**0.6752** $\pm$ 0.3922	0.6293 $\pm$ 0.4128 $\downarrow$	0.6315 $\pm$ 0.4122 $\downarrow$
PPV	**0.5524** $\pm$ 0.3684	0.5313 $\pm$ 0.3746	0.5221 $\pm$ 0.3773 $\downarrow$
Dice	**0.5765** $\pm$ 0.3579	0.5497 $\pm$ 0.3691 $\downarrow$	0.5439 $\pm$ 0.3706 $\downarrow$
AAD	**16267.9** $\pm$ 22450.7	16932.0 $\pm$ 21978.2	16708.8 $\pm$ 21573.4



(a) Original **(b)** Manual Annotations

(c) Expert 1 **(d)** Expert 2

(e) Automatic

Figure 6.11: Example of drusen detection using different segmentations to compute the features. The manual annotations of the left and right limits of drusen are represented by white and black columns, respectively. The columns marked in white and gray below the BM refer to the different drusen subclasses detected.

defined as the mean plus three standard deviations of the $dist_{IRPEDC\_BM}$ retrieved from a dataset of normal patients. In this case, the database does not have healthy patients, alternatively the threshold was learned through a 5-fold cross validation scheme that minimizes the cross entropy. The method of Dufour $et\ al.$ [6] was also evaluated. This method detects drusen through classification (section 6.1.1) and uses the segmentations of the IRPEDC and the BM computed with the method of Dufour $et\ al.$ [1] for retinas with drusen[19]. In [6], Dufour $et\ al.$ only report results of the AUC, thus it was assumed that this was also the metric used to learn

[19]This method was also included in the evaluation of the proposed method of segmentation of drusen limiting boundaries and it was described in more detail in section 5.2.4.

the hyperparameters of the RF classifier. Similar to the previous section, the proposed drusen detection method is based on the hybrid HMM-RF model using approach 3 (balanced) and 4 classes. Furthermore, the RF used to estimate the likelihood uses the LMFB and all the new features. These features were computed using the segmentations of the IRPEDC and the BM from the proposed segmentation method (chapter 5). Unlike the method of Dufour $et\ al.$ [6], the RF of the hybrid model minimizes the cross entropy (instead of the AUC), since this is a requirement for creating a hybrid HMM-RF model.

The method of Farsiu $et\ al.$ [133] uses thresholding to detect drusen, which is the most common approach used in the literature[20] (section 3.2). This approach has the advantage of being very simple, but it is also a limited one. Thresholding only considers the distance between the boundaries limiting drusen. As a consequence, the method of Farsiu $et\ al.$ [133] does not detect drusen or portions of drusen with height below the threshold (Fig. 6.12). The partial detection of drusen is reflected in the large number of small drusen detected and the low number of intermediate drusen detected by this method (Table 6.14). Additionally, using thresholding can misclassify regions of large RPEDC thickness that are not related with drusen.

To avoid the issues of thresholding, Dufour $et\ al.$ [6] use a classification approach that includes textural features of a neighborhood. This approach does not rely in the $dist_{IRPEDC\_BM}$, but rather on the morphology of the considered neighborhood (section 6.1.1). Comparing to the method of Farsiu $et\ al.$ [133], the method of Dufour $et\ al.$ [6] obtains a better column classification and a higher number of correctly detected drusen (Tables 6.13 and 6.14). On the other hand, the results in terms of drusen area measurements are similar for both methods (Table 6.13). This suggests that the segmentations of the Chiu $et\ al.$ [85] provide more accurate quantifications of drusen material than those of the method of Dufour $et\ al.$ [1] for retinas with drusen.

As previously mentioned, the proposed detection method improves upon the method of Dufour $et\ al.$ [6] by using new features, multi-label classification and the hybrid HMM-RF model (section 6.1.2). Furthermore, the segmentation of the proposed method (chapter 5) has a algorithm developed for handling drusen and GA and it uses local shape priors in the form of SHOPs to improve the segmentation of the IRPEDC. Compared to the other methods, the proposed method is more sensitive to drusen, but less sensitive to the the class of non-drusen (Fig. 6.12), as indicated by the sensitivity and specificity values of Table 6.13. The classification accuracy is worse for the proposed method, since the class of non-drusen is more pervasive than the class of drusen. These differences are mainly a consequence of balancing the importance of each class by training the RF in a balanced training set, as referred ealrier in section 6.3.2. Nevertheless, the proposed method produces a classification that is more similar to the manual markings than the remaining methods, as indicated by the higher dice coefficient (Table 6.13). These results are also reflected in the drusen area measurements, where the specificity of the proposed method is significantly lower and the sensitivity and dice are significantly higher than the other methods (Table 6.15). Moreover, the proposed method can correctly detect considerably more drusen than the remaining methods (Table 6.14). This is primarily related with the multi-label classification and the

[20]The thresholding approach was also compared to the method of Dufour $et\ al.$ [6] in appendix C. In this comparison, the segmentation of expert 1 was used for both methods.

Table 6.13: Results of column classification for the hybrid HMM-RF model and methods of Farsiu and Dufour. The values in bold refer to the best results of each row.

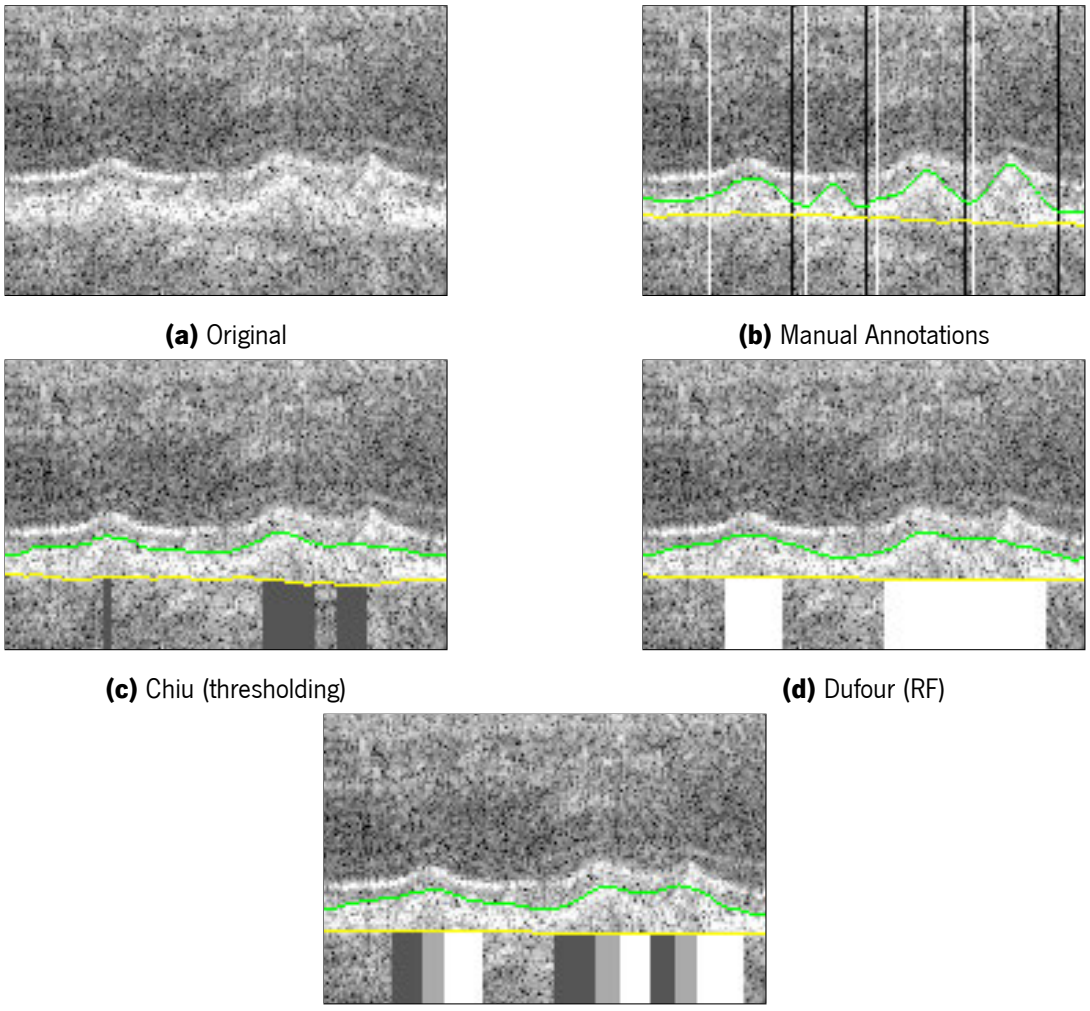
Segmentation	**Farsiu**	**Dufour**	**Proposed**
Sensitivity	0.5431	0.5801	**0.8343**
Specificity	0.9596	**0.9638**	0.8723
Accuracy	0.8781	**0.8888**	0.8649
Dice	0.6354	0.6711	**0.7072**

Table 6.14: Results of drusen detection for the hybrid HMM-RF model and methods of Farsiu and Dufour. The values in bold refer to the best results of each row.

	Segmentation	**Farsiu**	**Dufour**	**Proposed**
Detected drusen	**Small**	182	303	0
	Intermediate	95	108	23
	Large	311	378	867
	Total	588	789	890
Correctly detected drusen	**Small**	**1**	0	0
	Intermediate	10	13	**24**
	Large	59	126	**253**
	Total	70	139	**277**

Table 6.15: Results of drusen area measurements for the hybrid HMM-RF model and methods of Farsiu and Dufour. The AADs are presented in μm^2. Statistically significant results greater or lower than those that use the segmentation of expert 1 are represented by $\uparrow$ and $\downarrow$, respectively. Statistical significance was determined by Wilcoxon signed-ranked test (for paired data). The values in bold refer to the best results of each row.

Segmentation	**Farsiu**	**Dufour**	**Proposed**
Sensitivity	$0.4287 \pm 0.3573 \downarrow$	$0.4271 \pm 0.3682 \downarrow$	$\mathbf{0.6752} \pm 0.3922$
PPV	$\mathbf{0.5533} \pm 0.4410 \uparrow$	$0.5505 \pm 0.4457 \uparrow$	0.5524 ± 0.3684
Dice	$0.4500 \pm 0.3652 \downarrow$	$0.4554 \pm 0.3749 \downarrow$	$\mathbf{0.5765} \pm 0.3579$
AAD	14207.2 ± 15019.6	$\mathbf{13247.4} \pm 12967.6$	16267.9 ± 22450.7



(a) Original

(b) Manual Annotations

(c) Chiu (thresholding)

(d) Dufour (RF)

(e) Proposed (HMM-RF)

Figure 6.12: Example of drusen detection using the proposed method and the Farsiu and Dufour methods. The manual annotations of the left and right limits of drusen are represented by white and black columns, respectively. The columns marked in white and gray below the BM refer to the different drusen subclasses detected.

new features (section 6.3.1). The multi-label classification enables the individual detection of contiguous drusen, while the new features, more particularly the WCs, correlate well with the drusen subclasses, thus easing the classification (Fig. 6.12).

6.4 Summary

This chapter started by describing the proposed method for detecting drusen, which is based on the method of Dufour $et\ al.$ [6]. The latter method applies a RF classifier with features from the LMFB to classify each column as drusen or non-drusen. The proposed method extends this method by adding new features, performing multi-label classification and incorporating local information into the classification through a hybrid HMM-RF model.

Afterwards, it was evaluated the effect of the new features and the multi-label classification. The new features

improved upon the classification with just the LMFB outputs. The $dist_{IRPEDC\_BM}$ played an important role on correcting the left and right columns of drusen. As for the WCs, they contributed primarily to the improvement of individual detection of drusen when these occur in groups. The individual detection of drusen was only possible using a multi-label classification. This classification approach can also produce a more accurate classifier, since the use of multiple classes often favors the discrimination between drusen and non-drusen classes.

Thereafter, it was evaluated the incorporation of local information to the RF classifier by combining it with a HMM. The HMM-RF hybrid model attained a lower accuracy than the RF alone, nonetheless the local information improves the correct detection of drusen, largely by merging partially detected drusen. Moreover, the HMM-RF hybrid model also improves the dice coefficient of the classification and drusen area measurements. This implies that the results of this method are closer to the segmentations and markings of the expert graders than the results of the RF classifier.

Then, it was evaluated the robustness of the detection method to the use of other segmentations for computing features. Using the segmentation method proposed in chapter 5 generally leads to worse results, still these were comparable to the manual segmentations of both expert graders. This decrease in performance is primarily associated with the inherent presence of errors in automatic segmentations.

Finally, the proposed detection method was combined with the proposed segmentation method, described in chapter 5, and it was compared with two other methods present in the literature. The results indicate that the proposed method attains the highest degree of similarity to the manual delineations of drusen areas. Additionally, the proposed method also detects considerably more drusen correctly than the other methods.

Chapter 7

Conclusion

This work was focused in the quantification of drusen material in patients with intermediate AMD. The goal was to propose a method for quantifying automatically drusen material, instead of having to resort to the intervention of expert graders. Normally, the quantification is performed by using one method to segment the boundaries that limit drusen and another method for locating them. This approach was also followed in this work.

The proposed segmentation method was based on the MSS problem, which is solved with the MSF. This method proposes a new algorithm that is adapted to volumes of patients with intermediate AMD by taking into consideration the presence of lesions appearing in this stage of the disease, i.e., drusen and GA. Moreover, the proposed segmentation method also extends the MSS problem by integrating local shape priors. These priors are defined as SHOPs in the extended MSF. The SHOPs were introduced to improve the segmentation of the IRPEDC in drusen regions, which is frequently oversmoothed in these regions. The proposed segmentation method was evaluated against the methods of Chiu $et\ al.$ [85] and Dufour $et\ al.$ [1]. The method of Dufour $et\ al.$ for healthy retinas performed almost always considerably worse than the proposed method, demonstrating that the adaption of the algorithm to diseased retinas is important for accurately segmenting retinal boundaries. The method of Dufour $et\ al.$ for retinas with drusen performed better than the former, nonetheless, it performed considerably worse than the proposed method in extensive regions of GA, since this method did not account for lesions of this dimension. The proposed method also performed well in general against the method Chiu $et\ al.$ [85], which is a state of the art method that has its results publicly available. The effect of using SHOPs was also evaluated and their integration improved the segmentation of the IRPEDC. The improvement was mainly in drusen regions, but non-drusen regions were also improved, particularly situations where the IRPEDC was attracted to the IS-OS transitions or to light-to-dark transitions within the RPEDC. All of these results suggest that the new algorithm and the alterations to the MSF were effective and that the proposed method produces a reliable segmentation of the boundaries that limit drusen.

Drusen are frequently detected by thresholding the distance between boundaries limiting these lesions. This approach is simple, but it presents some problems, such as missing drusen or portions of drusen that are below the threshold. To avoid this issue, Dufour $et\ al.$ [6] proposed to detect drusen by classifying with a RF

each column as drusen or non-drusen. The features are local responses of the LMFB, which provide textural and morphological information of a neighborhood. In this work, it was proposed a detection method based on the work of Dufour $et\ al.$ [6]. The proposed method is capable of detecting drusen individually when these occur in clusters and it also integrates local information into the classification. For that, the proposed detection method adds the distance between the IRPEDC and the BM and the respective WCs as new features, performs a multi-label classification and uses a hybrid HMM-RF model instead of RF. The new features and the multi-label classification generally improved the results. The multi-label classification allowed the individual detection of contiguous drusen, which improved particularly the number of correctly detected drusen. Furthermore, the WCs were also important in this regard, since these features correlated well with drusen subclasses. The HMM-RF introduced local information to the classification. This information allowed the merging of several partially detected drusen, forbade physically impossible transitions between classes and smoothed the output of the classification, making it visually closer to the manual annotations.

The proposed segmentation and detection methods were combined to obtain a totally automatic quantification of drusen. This combination of methods was evaluated against the method of Dufour $et\ al.$ [6] and the method Farsiu $et\ al.$ [133]. The accuracy of the proposed method was lower than these two methods, nonetheless the results suggest that the output of proposed method is closer to the drusen delimitations of the experts graders. Furthermore, the proposed method correctly detects considerably more drusen than the other methods. These results suggest that the proposed method is capable of quantifying drusen material accurately. Still, this method could be further developed by improving several aspects, such as: the flattening, the definition of labelings of SHOPs, the E_{bound} term, the features of the drusen detection, among others. The flattening has the problem of not fully flatten part of the B-scans. Having a good flattening is important because it affects all the subsequent steps. If a perfect flattening could be attained, it would be expected that it would tighten the hard constraints limits, reducing the effect of noise. Furthermore, the BM would almost a be a straight line, which would ease its segmentation. The improvement of the quality of the segmentation would be reflected in the detection of drusen, since the errors of automatic segmentations deteriorate the detection results (subsection 6.3.3). As for the labelings of SHOPs, their definition still contains incorrect detections (subsection 5.3.2), so another an improved method for the defining the labelings could be beneficial. The E_{bound} is currently dependent on the gradient of the image, however, in some occasions the boundaries are not visible or lack contrast. In these situations, it may helpful to use other sources of information, as done in [75]. The proposed detection method has still room for improvement, for instance the classifier sometimes considers morphological alterations of the RPE as drusen, leading to false positives. Using features that could differentiate these situations could considerably improve the classification results. Nevertheless, this could be a difficult task, since there is not a fixed pattern for those morphological alterations. Thus, using features learned specifically for this problem could be a solution for this issue. By taking into account issues like these, the proposed method could be further developed to be part of a CAD system that quantifies drusen material to assess the evolution of AMD or to evaluate the effect of therapeutical procedures.

Appendices

Appendix A

Details of the segmentation algorithm

A.1 Parameter definition

The hyperparameters of the proposed algorithm were learned through a Bayesian optimization algorithm [125] (Table A.1). These parameters are mostly weights that control the soft smooth (λ) and interaction (α) constraints. The only exception is the weight that controls the SHOPs: p.

Table A.1: Minimum and maximum values of hyperparameters of all folds considered in the cross-validation. The hyperparameters are identified by the algorithm step because the boundaries can be segmented more than once.

Hyperparameter (step)		Min.	Max.	Hyperparameter (step)		Min.	Max.
λ_{ILM}	(step 3)	20	42	λ_{IRPEDC}	(step 6)	0	0
λ_{IRPEDC}	(step 4)	0	16	$\alpha_{IS\text{-}OS\_IRPEDC}$	(step 6)	13	17
λ_{BM}	(step 4)	228	296	λ_{IS-OS}	(step 8)	6	62
α_{IRPEDC_BM}	(step 4)	23	40	λ_{IRPEDC}	(step 8)	0	23
λ_{BM}	(step 5)	516	533	$\alpha_{IS\text{-}OS\_IRPEDC}$	(step 8)	18	26
$\lambda_{IS\text{-}OS}$	(step 6)	0	16	p	(step 8)	46	71

A.2 Analysis of processing times

The processing times presented in Table A.2 were acquired using a desktop computer with a processor AMD 8320 at 4.0GHz with 16GB of RAM running in a single CPU. For a single frame, the mean overall time spent in segmentation operations is 34.28 s. Lower mean processing times were observed when segmenting single boundaries than for segmentations with two boundaries (Table A.2) Moreover, similar processing times were observed by the simultaneous segmentation of IS-OS and IRPEDC with and without SHOPs (step 6 and 8). When

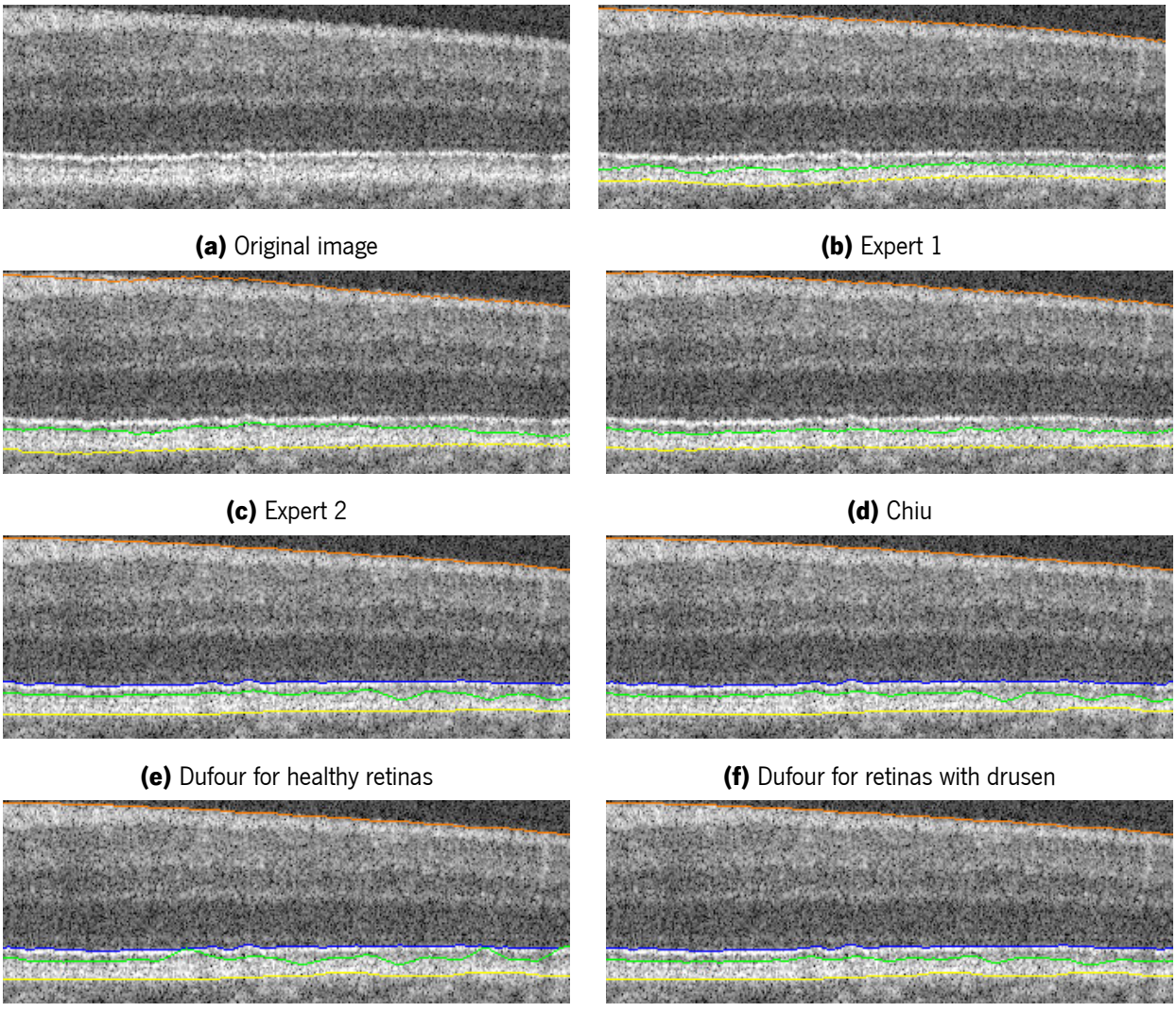
using SHOPs, the segmentation is delayed a mean of 0.55s (or 11%). This time increase is not drastic because only a reduced number of SHOPs was employed (step 8). Nonetheless, the integration of SHOPs requires two additional steps (step 7 and 8). These steps add 6.58 s of mean processing time, which represent 19% of the total segmentation time per frame.

Table A.2: Mean processing times ($\pm$ standard deviation) of each segmentation in seconds for a single CPU. The segmentations are identified by algorithm step because some occur more than once.

Boundaries	**(step)**	**Time**	**Boundaries**	**(step)**	**Time**
ILM; IRPEDC	(step 2)	12.81 $\pm$ 8.33 s	**IS-OS; IRPEDC**	(step 6)	5.09 $\pm$ 0.76 s
ILM	(step 3)	1.41 $\pm$ 0.15 s	**IRPEDC (intensity)**	(step 7)	0.50 $\pm$ 0.05 s
IRPEDC; BM	(step 4)	7.20 $\pm$ 0.98 s	**IRPEDC**	(step 7)	0.44 $\pm$ 0.04 s
BM	(step 5)	1.19 $\pm$ 0.25 s	**IS-OS; IRPEDC**	(step 8)	5.64 $\pm$ 0.70 s

Appendix B

Additional segmentation examples



(a) Original image

(b) Expert 1

(c) Expert 2

(d) Chiu

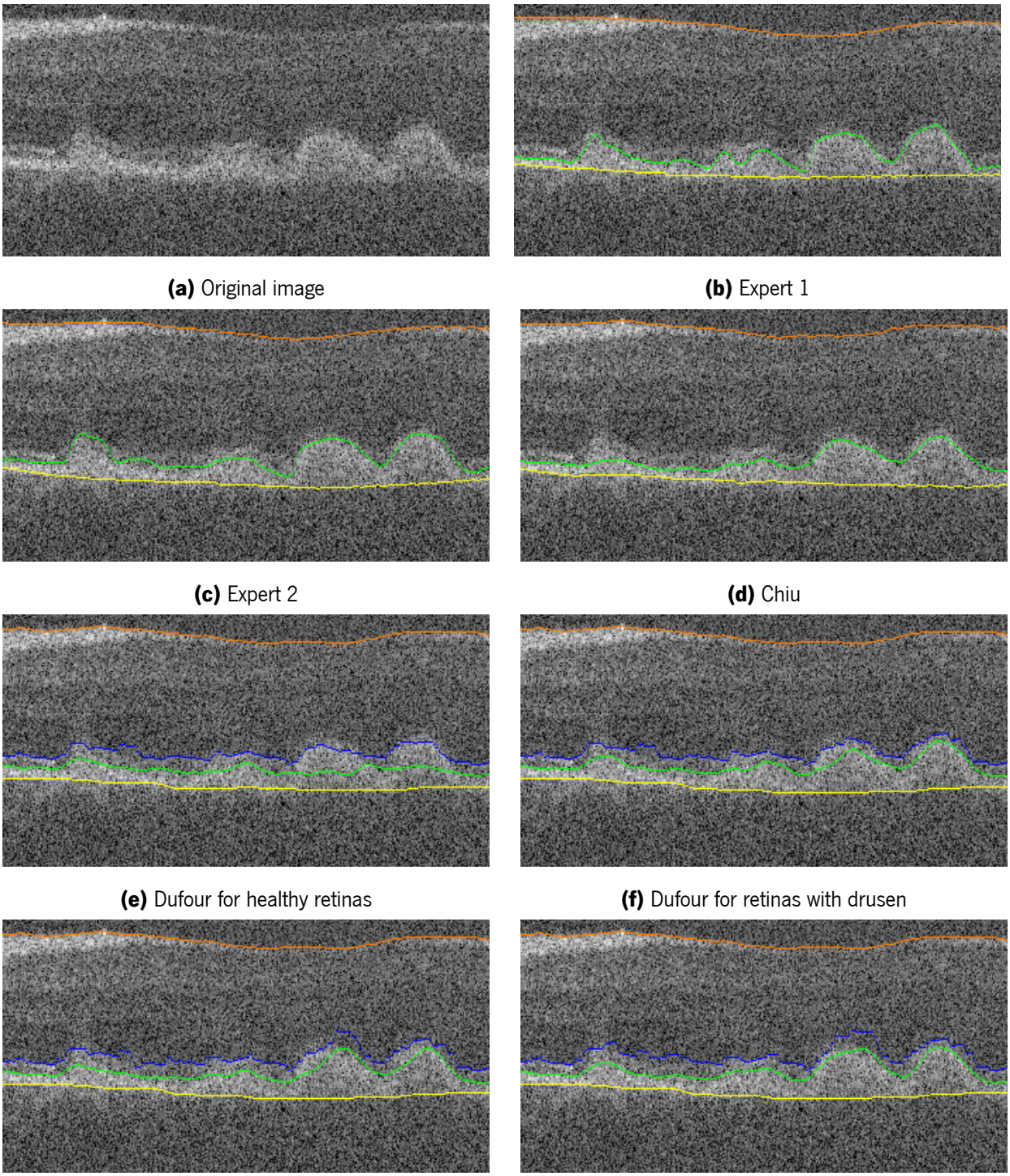
(e) Dufour for healthy retinas

(f) Dufour for retinas with drusen

(g) Proposed method wihtout SHOPs

(h) Proposed method

Figure B.1: Example of segmentations of several methods in a B-scan without drusen. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow)



(a) Original image

(b) Expert 1

(c) Expert 2

(d) Chiu

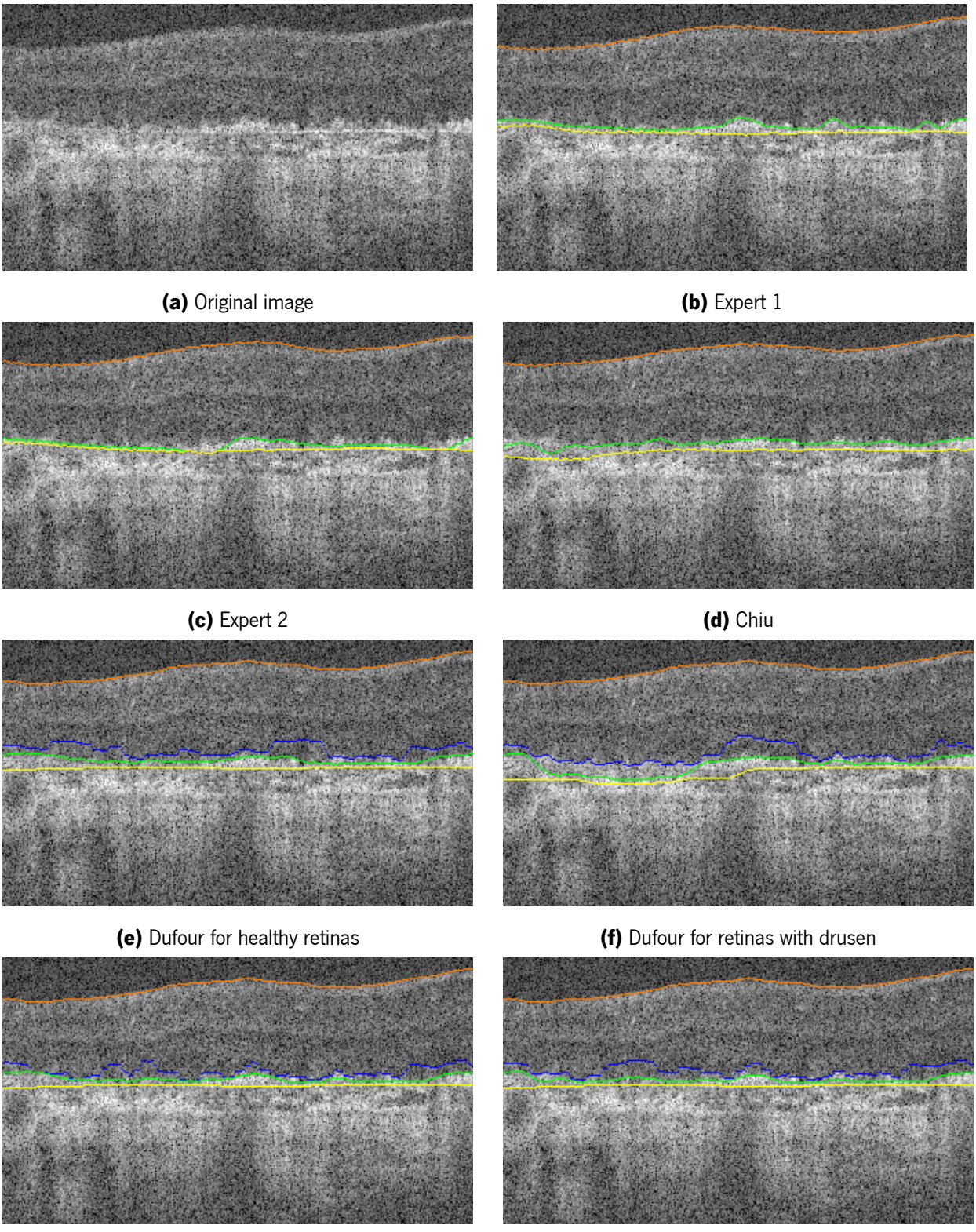
(e) Dufour for healthy retinas

(f) Dufour for retinas with drusen

(g) Proposed method wihtout SHOPs

(h) Proposed method

Figure B.2: Example of segmentations of several methods in a B-scan with drusen. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow)



(a) Original image

(b) Expert 1

(c) Expert 2

(d) Chiu

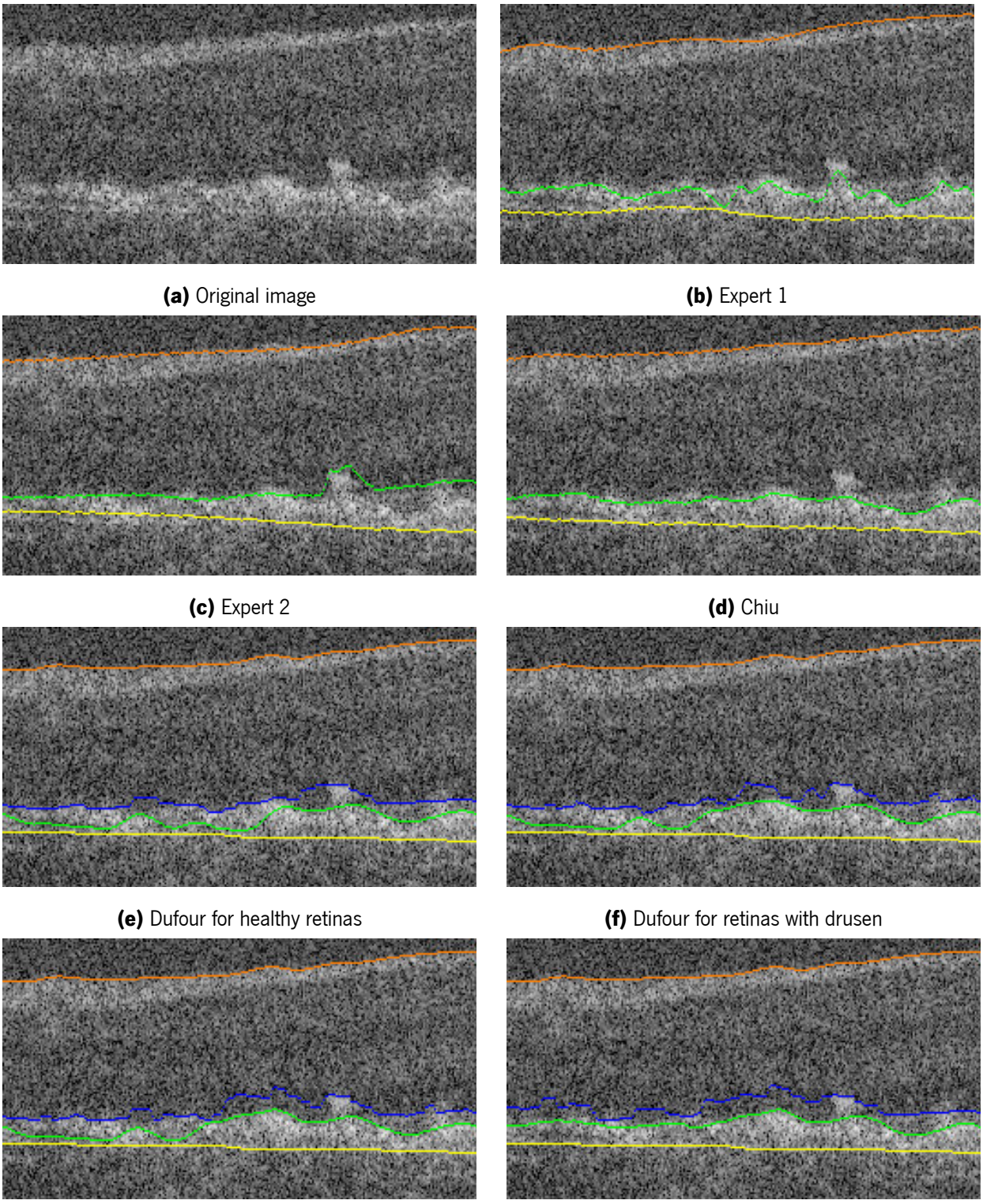
(e) Dufour for healthy retinas

(f) Dufour for retinas with drusen

(g) Proposed method wihtout SHOPs

(h) Proposed method

Figure B.3: Example of segmentations of several methods in a B-scan with extensive GA. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow)



(a) Original image

(b) Expert 1

(c) Expert 2

(d) Chiu

(e) Dufour for healthy retinas

(f) Dufour for retinas with drusen

(g) Proposed method wihtout SHOPs

(h) Proposed method

Figure B.4: Example of segmentations of several methods in a B-scan where the IRPEDC is attracted to transitions within the RPEDC. Depicted boundaries are: ILM (orange), IS-OS (blue), IRPEDC (green) and BM (yellow)

Appendix C

Evaluation of the detection method of Dufour $et\ al.$

Tables C.1, C.2, C.3 present the results of the method of Dufour $et\ al.$ [6]. The hyperparameters for the method Dufour $et\ al.$ were learned by defining the AUC or the cross entropy as loss function. In [6], Dufour $et\ al.$ only report results of the AUC, thus it was assumed that this was also the metric used to learn the hyperparameters. In this work, the AUC is replaced by the cross entropy, since it is necessary that the RF minimizes this metric to create a hybrid HMM-RF model.

The performance regarding column classification is similar for both the AUC and the cross entropy, even though most metrics suggest a slight improvement of performance for the cross entropy (Table C.1). The same behavior is supported by number of correctly detected drusen (Table C.2) and the drusen area measurements (Tables C.3).

Besides the method of Dufour $et\ al.$, tables C.1, C.2, C.3 also include the results of detecting drusen by thresholding. The results of this method are reported because it is the most common approach for locating drusen. The best threshold for detecting drusen was learned through a 5-fold cross validation scheme with cross entropy as evaluation metric. The overall performance of this approach is worse than the one of the method of Dufour $et\ al.$, as suggested by the column classification accuracy (Table C.1). This is mainly justified by a high number of false negatives, as indicated by the lower sensitivity (Table C.1). The high number of false negatives is related with the deficient detection of large drusen (Table C.2). The thresholding method is missing large drusen when these do not exhibit a maximum $dist_{IRPEDC\_BM}$ above the threshold[1] or when the maximum $dist_{IRPEDC\_BM}$ is above the threshold, but a considerable part of the drusen has the $dist_{IRPEDC\_BM}$ below the threshold (Fig. C.1).

[1]The dimension of drusen is defined according to the width and not the height.

Table C.1: Results of the method of Dufour $et\ al.$ and of the thresholding approach regarding the classification of columns. The values in bold refer to the best results of each row.

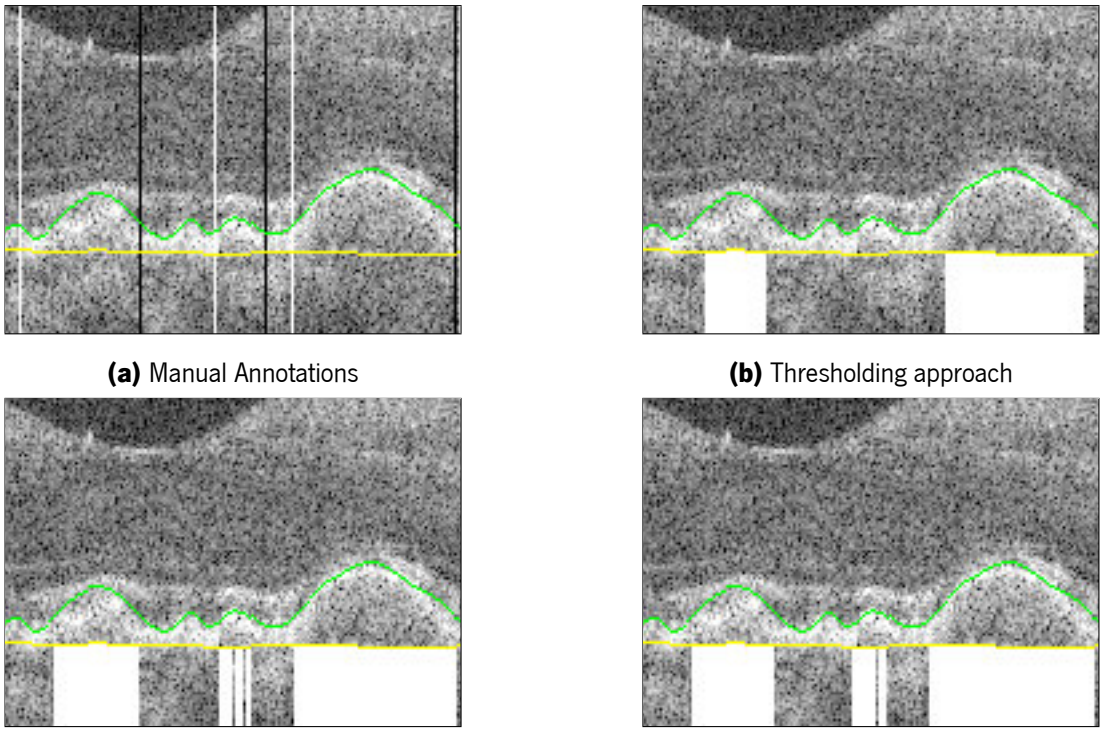
Method	**Dufour** $et\ al.$		**Thresholding**
Metric	**AUC**	**Cross entropy**	**Cross entropy**
Sensitivity	61.61 %	**62.58** %	53.28 %
Specificity	96.12 %	96.03 %	**96.41** %
Accuracy	89.37 %	**89.49 %**	87.98 %
AUC	91.29 %	**91.49 %**	–
Dice	69.40 %	**69.96 %**	63.41 %

Table C.2: Number of drusen detected by the method of Dufour $et\ al.$ and of the thresholding approach. For the correctly detected drusen, the largest values of each row are highlighted in bold.

Method		**Dufour** $et\ al.$		**Thresholding**	**Expert grader**
Metric		**AUC**	**Cross entropy**	**Cross entropy**	–
Detected drusen	Small	503	490	296	12
	Intermediate	113	109	242	149
	Large	369	381	317	646
	Total	985	980	855	807
Correctly detected drusen	Small	1	1	**3**	12
	Intermediate	17	21	**31**	149
	Large	141	**160**	78	646
	Total	159	**182**	112	807

Table C.3: Results of the method of Dufour $et\ al.$ and of the thresholding approach regarding drusen area measurements. The AADs are presented in μm^2. Statistically significant results greater or lower than those of the method of Dufour using the cross entropy are represented by $\uparrow$ and $\downarrow$, respectively. The values in bold refer to the best mean results of each row.

Method	**Dufour** $et\ al.$		**Thresholding**
Metric	**AUC**	**Cross entropy**	**Cross entropy**
Sensitivity	48.67 % $\pm$ 37.95 % $\downarrow$	**49.72 %** $\pm$ 37.81 %	46.02 % $\pm$ 32.56 % $\downarrow$
PPV	60.27 % $\pm$ 43.80 %	**61.32 %** $\pm$ 42.83 %	60.91 % $\pm$ 41.87 %
Dice	50.68 % $\pm$ 37.97 % $\downarrow$	**51.49 %** $\pm$ 37.66 %	49.23 % $\pm$ 33.82 % $\downarrow$
AAD	13534.33 $\pm$ 14407.73 $\uparrow$	**13145.53** $\pm$ 14170.50	14081.99 $\pm$ 12498.31 $\uparrow$



(a) Manual Annotations

(b) Thresholding approach

(c) RF using the AUC

(d) RF using the cross entropy

Figure C.1: Example of drusen detection using the thresholding approach and the method of Dufour $et\ al.$ with both the AUC and crossing entropy. The manual annotations of the left and right limits of drusen are represented by white and black columns respectively. The columns that each method considers as drusen were marked with white pixels below the BM.

References

[1] P. a. Dufour, L. Ceklic, H. Abdillahi, S. Schröder, S. De Dzanet, U. Wolf-Schnurrbusch, and J. Kowal, "Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints.," IEEE transactions on medical imaging, vol. 32, pp. 531–43, Mar. 2013.

[2] A. Yazdanpanah, G. Hamarneh, B. R. Smith, and M. V. Sarunic, "Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach.," IEEE transactions on medical imaging, vol. 30, pp. 484–96, Feb. 2011.

[3] M. Szkulmowski, M. Wojtkowski, B. Sikorski, T. Bajraszewski, V. J. Srinivasan, A. Szkulmowska, J. J. Kałuzny, J. G. Fujimoto, and A. Kowalczyk, "Analysis of posterior retinal layers in spectral optical coherence tomography images of the normal retina and retinal pathologies," Journal of biomedical optics, vol. 12, no. 4, pp. 041207–041207–11, 2007.

[4] S. Farsiu, S. J. Chiu, J. a. Izatt, and C. a. Toth, "Fast Detection and Segmentation of Drusen in Retinal Optical Coherence Tomography Images," Proc. SPIE, vol. 6844, pp. 68440D–68440D–12, Feb. 2008.

[5] Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. de Sisternes, and D. L. Rubin, "Automated drusen segmentation and quantification in SD-OCT images.," Med. Image Anal., vol. 17, pp. 1058–1072, Dec. 2013.

[6] P. Dufour, D. Z. S., W.-S. U., and K. J., "Classification of drusen positions in optical coherence tomography data from patients with age-related macular degeneration," in Int. Conf. on Pattern Recognition, pp. 2067–2070, 2012.

[7] T. Kauppi, Eye fundus image analysis for automatic detection of diabetic retinopathy. PhD thesis, 2010.

[8] B. Baumann, E. Gotzinger, M. Pircher, H. Sattmann, C. Schuutze, F. Schlanitz, C. Ahlers, U. Schmidt-Erfurth, and C. K. Hitzenberger, "Segmentation and quantification of retinal lesions in age-related macular degeneration using polarization-sensitive optical coherence tomography.," Journal of biomedical optics, vol. 15, no. 6, p. 061704, 2010.

[9] A. Wood, Retinal structure and function in age-related maculopathy. PhD thesis, Cardiff University, 2011.

[10] Acess Economics Pty Limited, "The Global Economic Cost of Visual Impairment," Tech. Rep. March, 2010.

[11] F. Soares, "Degenerescência Macular da Idade (DMI)." http://saude.sapo.pt/saude-em-familia/senior/degenerescencia-macular-da-idade-dmi.html, (accessed December 19, 2013).

[12] R. D. Jager, W. F. Mieler, J. W. Miller, K. B. Freund, J. M. Klancnik, L. A. Yannuzzi, B. Rosenthal, R. D. Jager, W. F. Mieler, and J. W. Miller, "Age-related macular degeneration," New England Journal of Medicine, vol. 358, pp. 2606–2617, June 2008.

[13] P. T. Johnson, G. P. Lewis, K. C. Talaga, M. N. Brown, P. J. Kappel, S. K. Fisher, D. H. Anderson, and L. V. Johnson, "Drusen-Associated Degeneration in the Retina," Investigative Ophthalmology & Visual Science, vol. 44, pp. 4481–4488, Oct. 2003.

[14] R. Haimovici, D. L. Gantz, S. Rumelt, T. F. Freddo, and D. M. Small, "The lipid composition of drusen, Bruch's membrane, and sclera by hot stage polarizing light microscopy," Investigative Ophthalmology & Visual Science, vol. 42, no. 7, pp. 1592–1599, 2001.

[15] N. Jain, S. Farsiu, A. a. Khanifar, S. Bearelly, R. T. Smith, J. a. Izatt, and C. a. Toth, "Quantitative comparison of drusen segmented on SD-OCT versus drusen delineated on color fundus photographs.," Investigative ophthalmology & visual science, vol. 51, pp. 4875–83, Oct. 2010.

[16] Age-Related Eye Disease Study Research Group, "The Age-Related Eye Disease Study severity scale for age-related macular degeneration: AREDS report No. 17.," Archives of Ophthalmology, vol. 123, no. 11, pp. 1484–1498, 2005.

[17] F. G. Schlanitz, C. Ahlers, S. Sacu, C. Schütze, M. Rodriguez, S. Schriefl, I. Golbaz, T. Spalek, G. Stock, and U. Schmidt-Erfurth, "Performance of drusen detection by spectral-domain optical coherence tomography.," Investigative ophthalmology & visual science, vol. 51, pp. 6715–21, Dec. 2010.

[18] D. Iwama, M. Hangai, S. Ooto, A. Sakamoto, H. Nakanishi, T. Fujimura, A. Domalpally, R. P. Danis, and N. Yoshimura, "Automated assessment of drusen using three-dimensional spectral-domain optical coherence tomography.," Investigative ophthalmology & visual science, vol. 53, pp. 1576–83, Mar. 2012.

[19] C. Duanggate and B. Uyyanonvara, "A Review of Automatic Drusen detection and segmentation from retinal images," in The 3rd International Symposium on Biomedical Engineering, vol. 1, pp. 222–225, 2008.

[20] L. Giancardo, Automated fundus images analysis techniques to screen retinal diseases in diabetic patients. PhD thesis, Unicersité de Bourgogne, 2011.

[21] J. Lim, Two-dimensional signal and image processing. New Jersey: Prentice Hall PTR, 1990.

[22] B. M. Koeppen and B. A. Stanton, Berne & Levy Physiology. Mosby Elsevier, 6th ed., 2008.

[23] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," IEEE reviews in biomedical engineering, vol. 3, pp. 169–208, Jan. 2010.

[24] Lee Ann Remington, Clinical anatomy and physiology of the Visual System. Elsevier Butterworth Heinemann, 3rd ed., 2012.

[25] The Angiogenesis Foundation, "The science of AMD." http://www.scienceofamd.org/learn/, (accessed December 5, 2013).

[26] Donnelly Optical, "What is Age-Related Macular Degeneration." http://www.donnellyoptical.com/what-is-age-related-macular-degeneration/, 2013 (accessed December 5, 2013).

[27] Age-Related Eye Disease Study Research Group, "The Age-Related Eye Disease Study System for Classifying Age-related Macular Degeneration from Stereoscopic Color Fundus Photographs: the Age-Related Eye Disease Study Report Number 6," American journal of ophthalmology, vol. 132, no. 5, pp. 668–681, 2001.

[28] F. G. Schlanitz, B. Baumann, M. Kundi, S. Sacu, M. Baratsits, U. Scheschy, A. Shahlaee, T. J. Mittermüller, A. Montuoro, P. Roberts, M. Pircher, C. K. Hitzenberger, and U. Schmidt-Erfurth, "Drusen volume development over time and its relevance to the course of age-related macular degeneration," Br. J. Ophthalmol., vol. 0, pp. 1–6, 2016.

[29] Age-Related Eye Disease Study Research Group, "Risk Factors Associated with Age-Related Macular Degeneration," Ophthalmology, vol. 107, no. 12, pp. 2224–2232, 2000.

[30] U.S. Department of health and human services, "Age-Related Macular Degeneration: What you should know." 2009.

[31] S. Zayit-soudry, I. Moroz, and A. Loewenstein, "Retinal Pigment Epithelial Detachment," Survey of ophthalmology, vol. 52, no. 3, pp. 227–243, 2007.

[32] Photospectra Health Sciences, "Age Related Macular Degeneration." http://www.photospectra.ca/aboutAMD.php, 2011 (accessed December 10, 2013).

[33] The Age-Related Eye Disease Study 2 Research Group, "Lutein + Zeaxanthin and Omega-3 Fatty Acids for Age-Related Macular Degeneration," Journal of the American Medical Association, vol. 309, no. 19, pp. 2005–2015, 2013.

[34] F. H. Hengerer, P. Artal, T. Kohnen, and I. Conrad-hengerer, "Initial Clinical Results of a New Telescopic IOL Implanted in Patients With Dry Age- Related Macular Degeneration," Journal of Refractive Surgery, vol. 31, no. 3, pp. 158–162, 2015.

[35] London Eye Hospital Pharma, "Is iolAMD suitable for you?." http://iolamd.com/right-for-me.php, 2014 (accessed October 21, 2016).

[36] Wannan, Olivia and Derbyshire, Laurence, "Dry AMD treatment's 'major achievment'." https://www.aop.org.uk/ot/science-and-vision/research/2016/07/27/dry-amd-treatments-major-achievement, 2016 (accessed October 21, 2016).

[37] F. G. Holz, S. Schmitz-valckenberg, and M. Fleckenstein, "Recent developments in the treatment of age-related macular degeneration," The Journal of Clinical Investigation, vol. 124, no. 4, pp. 1430–1438, 2014.

[38] W. Drexler and J. Fujimoto, Optical coherence tomography: technology and applications. Springer, 2008.

[39] G. Gregori, F. Wang, P. J. Rosenfeld, Z. Yehoshua, N. Z. Gregori, B. J. Lujan, C. a. Puliafito, and W. J. Feuer, "Spectral domain optical coherence tomography imaging of drusen in nonexudative age-related macular degeneration.," Ophthalmology, vol. 118, pp. 1373–9, July 2011.

[40] Y. Y. Huang, A. C. H. Chen, and M. Hamblin, "Low-level laser therapy: an emerging clinical paradigm," SPIE Newsroom, pp. 2–4, 2009.

[41] S. L. Jacques, "Optical properties of biological tissues: a review," Phys. Med. Biol., vol. 58, no. 11, pp. R37–R61, 2013.

[42] Renishaw, "Interferometry explained." http://www.renishaw.com/en/interferometry-explained--7854, 2001-2016 (accessed November 8, 2016).

[43] W. Drexler and J. G. Fujimoto, Optical Coherence Tomography. Springer International Publishing, 2nd ed., 2015.

[44] Z. Yaqoob, J. Wu, and C. Yang, "Spectral domain optical coherence tomography: a better OCT imaging strategy.," BioTechniques, vol. 39, pp. S6—13, Dec. 2005.

[45] N. Hodgson and H. Weber, Laser Resonators and Beam Propagation: Fundamentals, Advanced Concepts, Applications. Springer Series in Optical Sciences, Springer Berlin Heidelberg, 2005.

[46] M. R. Garcia, "The Principles Behind OCT." https://wiki.engr.illinois.edu/display/BIOE414/The+Principles+Behind+OCT, (accessed October 30, 2016).

[47] E. Bousi and C. Pitris, "Lateral resolution improvement in Optical Coherence Tomography (OCT) images," in Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering, no. November, pp. 11–13, 2012.

[48] A. Ritscher, "„Pressure dependence of the Sn-P phase diagram and investigations in the ternary system Ni-P-Sn"," 2012.

[49] B. Smith, "What Is the Depth of Focus of a Microscope?." http://www.ehow.com/about_5148199_depth-focus-microscope.html, (accessed November 1, 2013).

[50] F. Shi, X. Chen, H. Zhao, W. Zhu, D. Xiang, E. Gao, M. Sonka, and H. Chen, "Automated 3-D Retinal Layer Segmentation of Macular Optical Coherence Tomography Images with Serous Pigment Epithelial Detachments.," IEEE transactions on medical imaging, vol. 34, pp. 441–452, Oct. 2014.

[51] D. C. Debuc, "A review of algorithms for segmentation of retinal image data using optical coherence tomography," in Image Segmentation (P.-G. Ho, ed.), ch. 2, pp. 15–54, InTech, 2011.

[52] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map.," Medical image analysis, vol. 17, pp. 907–28, Dec. 2013.

[53] M. Mayer, J. Hornegger, C. Mardin, and R. Tornow, "Retinal Nerve Fiber Layer Segmentation on FD-OCT Scans of Normal Subjects and Glaucoma Patients.," Biomed. Opt. Express, vol. 1, pp. 1358–1383, Jan. 2010.

[54] D. Koozekanani, K. Boyer, and C. Roberts, "Retinal Thickness Measurements From Optical Coherence Tomography Using a Markov Boundary Model," IEEE Trans. Med. Imag., vol. 20, no. 9, pp. 900–916, 2001.

[55] V. J. Srinivasan, B. K. Monson, M. Wojtkowski, R. A. Bilonick, I. Gorczynska, R. Chen, J. S. Duker, J. S. Schuman, and J. G. Fujimoto, "Characterization of Outer Retinal Morphology with High-Speed, Ultrahigh-Resolution Optical Coherence Tomography," Investigative ophthalmology & visual science, vol. 49, pp. 1571–9, apr 2008.

[56] D. Cabrera Fernández, H. M. Salinas, and C. a. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images.," Optics express, vol. 13, pp. 10200–16, Dec. 2005.

[57] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, "Macular Segmentation with Optical Coherence Tomography," Investigative Ophthalmology & Visual Science, vol. 46, no. 6, pp. 2012–2017, 2005.

[58] O. Tan, A. T.-h. Lu, R. Varma, and D. Huang, "Mapping of Macular Substructures with Optical Coherence Tomography for Glaucoma Diagnosis," Ophthalmology, vol. 115, no. 6, pp. 949–956, 2007.

[59] T. F. Chan and L. A. Vese, "Active Contours Without Edges," IEEE transactions on image processing, vol. 10, no. 2, pp. 266–277, 2001.

[60] A. Carass, A. Lang, M. Hauser, P. A. Calabresi, H. S. Ying, and J. L. Prince, "Multiple-object geometric deformable model for segmentation of macular OCT," Biomedical Optics Express, vol. 5, no. 4, pp. 1062–1074, 2014.

[61] J. A. Bogovic, J. L. Prince, and P.-l. Bazin, "A multiple object geometric deformable model for image segmentation," Computer Vision and Image Understanding, vol. 117, no. 2, pp. 145–157, 2013.

[62] A. Lang, A. Carass, M. Hauser, E. S. Sotirchos, P. a. Calabresi, H. S. Ying, and J. L. Prince, "Retinal layer segmentation of macular OCT images using boundary classification.," Biomedical optics express, vol. 4, pp. 1133–52, July 2013.

[63] J. Novosel, G. Thepass, H. G. Lemij, J. F. D. Boer, K. A. Vermeer, and L. J. V. Vliet, "Loosely coupled level sets for simultaneous 3D retinal layer segmentation in optical coherence tomography," Medical Image Analysis, vol. 26, no. 1, pp. 146–158, 2015.

[64] K. a. Vermeer, J. Mo, J. J. a. Weda, H. G. Lemij, and J. F. de Boer, "Depth-resolved model-based reconstruction of attenuation coefficients in optical coherence tomography," Biomedical Optics Express, vol. 5, no. 1, pp. 322–337, 2013.

[65] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," Numerisch Mathematik, vol. 1, pp. 269–271, 1959.

[66] F. Rossant, I. Bloch, I. Ghorbel, and M. Paques, "Parallel Double Snakes. Application to the segmentation of retinal layers in 2D-OCT for pathological subjects," Pattern Recognition, vol. 48, no. 12, pp. 3857–3870, 2015.

[67] N. V. Rangaswamy, H. M. Patel, K. G. Locke, D. C. Hood, and D. G. Birch, "A Comparison of Visual Field Sensitivity to Photoreceptor Thickness in Retinitis Pigmentosa," Investigative Ophthalmology & Visual Science, vol. 51, no. 8, pp. 4213–4219, 2010.

[68] H. Zhu, D. P. Crabb, P. G. Schlottmann, T. Ho, and D. F. Garway-Heath, "FloatingCanvas: quantification of 3D retinal structures from spectral-domain optical coherence tomography.," Optics express, vol. 18, pp. 24595–610, Nov. 2010.

[69] V. Kajić, B. Povazay, B. Hermann, B. Hofer, D. Marshall, P. L. Rosin, and W. Drexler, "Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis," Optics express, vol. 18, pp. 14730–44, jul 2010.

[70] F. Rathke, S. Schmidt, and C. Schnörr, "Probabilistic intra-retinal layer segmentation in 3-D OCT images using global shape regularization.," Medical image analysis, vol. 18, pp. 781–94, July 2014.

[71] K. A. Vermeer, J. van der Schoot, H. G. Lemij, and J. F. de Boer, "Automated segmentation by pixel classification of retinal layers in ophthalmic OCT images.," Biomedical optics express, vol. 2, pp. 1743–56, jun 2011.

[72] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map," Medical Image Analysis, vol. 17, no. 8, pp. 907–928, 2013.

[73] J. Canny, "A Computational Approach to Edge Detectio," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679–698, 1986.

[74] K. Li, X. Wu, D. Z. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images–a graph-theoretic approach.," IEEE transactions on pattern analysis and machine intelligence, vol. 28, pp. 119–34, Jan. 2006.

[75] B. J. Antony, M. D. Abràmoff, M. M. Harper, W. Jeong, E. H. Sohn, Y. H. Kwon, R. Kardon, and M. K. Garvin, "A combined machine-learning and graph-based framework for the segmentation of retinal surfaces in SD-OCT volumes.," Biomedical optics express, vol. 4, pp. 2712–28, Jan. 2013.

[76] M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, "Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images.," IEEE transactions on medical imaging, vol. 28, pp. 1436–47, Sept. 2009.

[77] S. J. Chiu, X. T. Li, P. Nicholas, C. a. Toth, J. a. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation.," Opt. express, vol. 18, pp. 19413–28, Aug. 2010.

[78] Q. Yang, C. a. Reisman, Z. Wang, Y. Fukuma, M. Hangai, N. Yoshimura, A. Tomidokoro, M. Araie, A. S. Raza, D. C. Hood, and K. Chan, "Automated layer segmentation of macular OCT images using dual-scale gradient information.," Optics express, vol. 18, pp. 21293–307, Sept. 2010.

[79] J. Tian, B. Varga, G. M. Somfai, W.-h. Lee, W. E. Smiddy, and D. C. Debuc, "Real-Time Automatic Segmentation of Optical Coherence Tomography Volume Data of the Macular Region," vol. D, pp. 1–20, 2015.

[80] Q. Song, X. Wu, and Y. Liu, "Simultaneous searching of globally optimal interacting surfaces with shape priors," in IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2879–2886, 2010.

[81] F. A. Folgar, E. L. Yuan, M. B. Sevilla, S. J. Chiu, S. Farsiu, E. Y. Chew, and C. A. Toth, "Drusen Volume and Retinal Pigment Epithelium Abnormal Thinning Volume Predict 2-Year Progression of Age-Related Macular Degeneration," Ophthalmology, vol. 123, no. 1, pp. 39–50e1, 2016.

[82] K. Yi, M. Mujat, B. H. Park, W. Sun, J. W. Miller, J. M. Seddon, L. H. Young, J. F. de Boer, and T. C. Chen, "Spectral domain optical coherence tomography for quantitative evaluation of drusen and associated structural changes in non-neovascular age-related macular degeneration.," The British journal of ophthalmology, vol. 93, pp. 176–81, Feb. 2009.

[83] M. Mujat, R. Chan, B. Cense, B. Park, C. Joo, T. Akkin, T. Chen, and J. de Boer, "Retinal nerve fiber layer thickness map determined from optical coherence tomography images.," Optics express, vol. 13, pp. 9480–91, Nov. 2005.

[84] R. W. Schafer, "What is a savitzky-golay filter?," IEEE Signal Processing Magazine, no. July, pp. 111–117, 2011.

[85] S. J. Chiu, J. A. Izatt, R. V. O. Connell, K. P. Winter, C. A. Toth, and S. Farsiu, "Validated Automatic Segmentation of AMD Pathology Including Drusen and Geographic Atrophy in SD-OCT Images," Investigative Ophthalmology & Visual Science, vol. 53, no. 1, pp. 53–61, 2012.

[86] T. Leung and J. Malik, "Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons," vol. 43, no. 1, pp. 29–44, 2001.

[87] J. Novosel, Z. Wang, H. D. Jong, K. A. Vermeer, and L. J. V. Vliet, "Loosely coupled level sets for retinal layers and drusen segmentation in subjects with dry age-related macular degeneration," in Medical Imaging 2016: Image Processing, vol. 9784, pp. 97842P–97842P–7, 2016.

[88] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 1124–1137, Sept. 2004.

[89] J.-C. Picard, "Maximal Closure of a Graph and Applications to Combinatorial Problems," Management Science, vol. 22, no. 11, pp. 1268–1272, 1976.

[90] D. S. Hochbaum and M. Queyranne, "Minimizing a convex cost closure set," SIAM Journal on Discrete Mathematics, vol. 16, no. 2, pp. 192–207, 2003.

[91] D. S. Hochbaum, "Lecture Notes for IEOR 269 - Integer Programming and Combinatorial Optimization." http://www.ieor.berkeley.edu/~hochbaum/files/ieor269-2010.pdf, 2010.

[92] City Sights, "Kennecott Copper Mine - Things to do in Salt Lake city." https://toursofutah.com/things-to-do-in-salt-lake-city/2014/5/29/kennecott-copper-mine-things-to-do-in-salt-lake-city, 2014 (accessed January 8, 2018).

[93] O. Danek, Graph Cut Based Image Segmentation in Fluorescence Microscopy. PhD thesis, Masaryk University, 2012.

[94] D. S. Hochbaum, "The Pseudoflow Algorithm: A New Algorithm for the Maximum-Flow Problem," Operations Research, vol. 56, no. 4, pp. 992–1009, 2008.

[95] M. K. Garvin, Automated 3-D segmentation and analysis of retinal optical coherence tomography images. PhD thesis, University of Iowa, 2008.

[96] M. Wels, G. Carneiro, A. Aplas, M. Huber, J. Hornegger, and D. Comaniciu, "A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI," in Med. Image Comput. Comput. Assist. Interv., pp. 67–75, 2008.

[97] S. Bagon, Discrete Energy Minimization, beyond Submodularity: Applications and Approximations. PhD thesis, Weizmann Institute of Science, 2012.

[98] N. Komodakis and N. Paragios, "Beyond pairwise energies: Efficient optimization for higher-order MRFs," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2985–2992, June 2009.

[99] P. Kohli, M. P. Kumar, and P. H. S. Torr, "P3 & Beyond: Solving Energies with Higher Order Cliques," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, Ieee, jun 2007.

[100] C. Rother and P. Kohli, "Minimizing sparse higher order energy functions of discrete variables," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1382–1389, Ieee, June 2009.

[101] S. Roth and M. J. Black, "Fields of Experts," International Journal of Computer Vision, vol. 82, pp. 205–229, Jan. 2009.

[102] A. Osokin and D. Vetrov, "Submodular Relaxation for MRFs with High-Order Potentials," in International Conference on Computer Vision, vol. 7585, 2012.

[103] C. Arora, S. Banerjee, P. Kalra, and S. N. Maheshwari, "Generic cuts: an efficient algorithm for optimal inference in Higher Order MRF-MAP," in Proc. European Conference on Computer Vision, (Florence), pp. 17–30, Springer-Verlag, 2012.

[104] H. Ishikawa, "Higher-order clique reduction in binary graph cut," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 2993–3000, jun 2009.

[105] A. Fix, A. Gruber, E. Boros, and R. Zabih, "A graph cut algorithm for higher-order Markov Random Fields," in International Conference on Computer Vision, pp. 1020–1027, nov 2011.

[106] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer, "Optimizing Binary MRFs via Extended Roof Duality," 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, June 2007.

[107] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[108] A. Criminisi and J. Shotton, eds., Decision Forests for Computer Vision and Medical Image Analysis. Springer, 2013.

[109] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning," tech. rep., Microsoft Research, 2011.

[110] Zoubin Ghahramani, "An Introduction to Hidden Markov Models and Bayesian Networks," Journal of Pattern Recognition and Artificial Intelligence, vol. 15, no. 1, pp. 9–42, 2001.

[111] Lawrence R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.

[112] D. Forney, "The Viterbi Algorithm," Proceedings of the IEEE, vol. 61, no. 3, pp. 268–278, 1973.

[113] F. Grézl and P. Fousek, "Optimizing bottle-neck features for LVCSR," in IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4729–4732, 2008.

[114] H. Hermansky, D. P. W. Ellis, and S. Sharma, "Tandem connectionist feature extraction for conventional HMM systems," in IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1635–1638 vol.3, 2000.

[115] S. Renals, N. Morgan, H. Bourlard, M. Cohen, and H. Franco, "Connectionist probability estimators in HMM speech recognition," IEEE Transactions on Speech and Audio Processing, vol. 2, no. 1, pp. 161–174, 1994.

[116] P. Pujol, S. Pol, C. Nadeu, A. Hagen, and H. Bourlard, "Comparison and Combination of Features in a Hybrid HMM/MLP and a HMM/GMM Speech Recognition System," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 1, pp. 14–22, 2005.

[117] L. Tóth and A. Kocsor, "Training HMM/ANN Hybrid Speech Recognizers by Probabilistic Sampling," in International Conference on Artificial Neural Networks, pp. 597–603, 2005.

[118] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using Context-Dependent Deep Neural Networks," in Conference of the International Speech Communication Association, INTERSPEECH, pp. 437–440, 2011.

[119] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Large vocabulary continuous speech recognition with context-dependent DBN-HMMs," IEEE International Conference on Acoustics, Speech and Signal Processin, pp. 4688–4691, 2011.

[120] D. Yu and L. Deng, Deep Neural Network-Hidden Markov Model Hybrid Systems. 2015.

[121] M. Gales and S. Young, "The Application of Hidden Markov Models in Speech Recognition," Foundations and Trends in Signal Processing, vol. 1, no. 3, pp. 195–304, 2008.

[122] M. D. Richard and R. P. Lippmann, "Neural Network Classifiers Estimate Bayesian a posteriori Probabilities," Neural Computation, vol. 3, no. 4, pp. 461–483, 1991.

[123] V. Punyakanok and D. Roth, "The Use of Classifiers in Sequential Inference," in Advances in Neural Information Processing Systems 13, pp. 995–1001, MIT Press, 2000.

[124] A. I. García-Moral, R. Solera-Ureña, C. Peláez-Moreno, and F. Díaz-de María, "Data balancing for efficient training of Hybrid ANN/HMM Automatic Speech Recognition systems," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 3, pp. 468–481, 2011.

[125] J. M. Hernández-Lobato, M. A. Gelbart, M. W. Hoffman, R. P. Adams, and Z. Ghahramani, "Predictive Entropy Search for Bayesian Optimization with Unknown Constraints," in Proceedings of the 32nd Int. Conf. on Mach. Learning, pp. 1699–1707, JMLR.org, 2015.

[126] L. Yin, R. Yang, and M. Gabbouj, "Weighted Median Filters: A Tutorial," IEEE Transactions on Circuits and Systems, vol. 43, no. 3, pp. 157–192, 1996.

[127] P. Ndajah, H. Kikuchi, M. Yukawa, H. Watanabe, and S. Muramatsu, "An investigation on the quality of denoised images," International Journal of Circuits, Systems and Signal Processing, vol. 5, no. 4, pp. 423–434, 2011.

[128] P. P. Srinivasan, S. J. Heflin, J. a. Izatt, V. Y. Arshavsky, and S. Farsiu, "Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology," Biomed. Opt. Express, vol. 5, p. 348, Jan. 2014.

[129] L. D. Sisternes, N. Simon, R. Tibshirani, T. Leng, and D. L. Rubin, "Quantitative SD-OCT Imaging Biomarkers as Indicators of Age-Related Macular Degeneration Progression," Invest. Ophthalmol. Vis. Sci., vol. 55, no. 11, pp. 7093–7103, 2014.

[130] Visual Geometry Group – University of Oxford, "Texture Classification." http://www.robots.ox.ac.uk/~vgg/research/texclass/filters.html, 2007 (accessed February 18, 2016).

[131] N. P. Hughes, L. Tarassenko, and S. J. Roberts, "Markov Models for Automated ECG Interval Analysis," in NIPS, pp. 611–618, 2004.

[132] W. Sun, H. Zhang, and A. Palazoglu, "Prediction of 8 h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models," Atmospheric Environment, vol. 81, pp. 199–208, 2013.

[133] S. Farsiu, S. J. Chiu, R. V. O'Connell, F. A. Folgar, E. Yuan, J. A. Izatt, and C. A. Toth, "Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography.," Ophthalmology, 2014.