

MODELING THE INTENSITY PROFILES OF
RETINAL VESSELS: AN EXTENDED HERMITE
MODEL AND ITS VALIDATION

RELATORE: Ch.mo Prof. Alfredo Ruggeri

CORRELATORE: Prof. Emanuele Trucco

LAUREANDA: Paola Nicoli

A.A. 2009-2010



UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
TESI DI LAUREA SPECIALISTICA IN BIOINGEGNERIA

MODELING THE INTENSITY
PROFILES OF RETINAL VESSELS:
AN EXTENDED HERMITE MODEL
AND ITS VALIDATION

RELATORE: Ch.mo Prof. Alfredo Ruggeri

CORRELATORE: Prof. Emanuele Trucco

LAUREANDA: *Paola Nicoli*

Padova, 13 Luglio 2010

Contents

Introduction	2
Rationale	2
Research goals and approach	4
Thesis outline	5
1 Clinical background	7
1.1 Introduction	7
1.2 The retina	7
1.2.1 Bloody supply to the retina	8
1.3 Diabetes Mellitus	9
1.3.1 Epidemiology	12
1.4 Diabetic Retinopathy	12
1.4.1 The Pathogenesis of Diabetic Retinopathy	13
1.5 Retinal Vessel Structure	14
1.6 Conclusions	16
2 Previous work for measuring retinal vessel profiles	17
2.1 Introduction	17
2.2 Introduction to diameter measurement algorithms	18
2.3 Models for fitting retinal vessel profiles	19
2.3.1 1-D Gaussian Model: Zhou and al.	19
2.3.2 2-D Gaussian Model: Lowell, Hunter et al.	21
2.3.3 Multiresolution Hermite Model (MHM): Wang et al.	22
2.4 Algorithms for measuring vascular diameters	27
2.4.1 FWHM (<i>F</i> ull <i>W</i> idth <i>H</i> alf <i>M</i> aximum): Brinchmann, Hansen et al.	27
2.4.2 Kick points: Rassam et al.	29
2.4.3 SLRF method (Sliding Linear Regression Filter): Chapman et al.	31
2.5 Conclusions	32

3	Extension of the Multiresolution Hermite Model	34
3.1	Introduction	34
3.2	Why we decided to extend the MHM	34
3.3	Experimental model setup	35
3.3.1	Summary of the Extended Hermite Model	39
3.4	Introduction to PSO	40
3.4.1	Particle Swarm Optimization, an optimization technique	40
3.4.2	PSO algorithm	41
3.5	Conclusion	43
4	Implementation of Extended MHM and PSO	44
4.1	Introduction	44
4.2	Extraction of intensity retinal vessel profile	44
4.3	Adaption of Extended MHM parameters	47
4.3.1	Variation of the parameter a and σ	47
4.3.2	Variation of the parameter δ and σ	48
4.3.3	Variation of the parameter m and σ	48
4.3.4	Variation of the parameter a and δ	49
4.3.5	Variation of the parameter p and σ	49
4.4	Adapting PSO (<i>Particle Swarm Optimization</i>) algorithm	51
4.4.1	Error function: Weight function	51
4.4.2	Initialization of the population of particles	51
4.4.3	Identification of the number of iterations	54
4.4.4	Values returned by PSO	56
4.5	Conclusions	56
5	Implementation of algorithms for measuring vessel diameter	60
5.1	Introduction	60
5.2	Implementation of flex points technique	60
5.3	Implementation of FWHM algorithm	61
5.3.1	Search of maximum values	63
5.3.2	Search of minimum values	64
5.4	Conclusions	67
6	Validation of the Extended MHM	68
6.1	Introduction	68
6.2	Testing accuracy and robustness of the model	68
6.3	Generating the Ground Truth	73

Introduction

Rationale

Diabetes is an increasingly common and serious condition often leading to multiple complications, largely as a consequence of macrovascular and microvascular diseases.

There's an urgent need to understand the biological mechanism involved in order to make a prior identification of this disease and improve targeting of more effective and appropriate clinical interventions.

This need arises from the numerous eye and systemic diseases that affect vessels of retina and can be diagnosed by detecting change in retina vessels.

There are multiple eye diseases that affect the vessels of retina, which can cause changes to existing vessels, in width, color and path; and can also cause growth of new vessels (neovascularization) [1].

These diseases are increasing enormously in recent years; statistics of blinding disease in the US provided by the University of Washington (*Department of Ophthalmology*) show that:

- 80 million people suffer from potentially blinding eye disease
- 1,100,000 people are legally blind
- 7 million diabetics suffer from diabetic retinopathy
- 700,000 diabetics are presently at risk of blindness
- 25,000 new cases of blindness are caused yearly by complications of diabetes
- 650,000 people are hospitalized each year for eye injury or disease

A study commencing in 1997 created a very large case control collection to investigate the genetics of type 2 Diabetes (TD2): patients with diabetes are offered regular retinal screening to assess and manage the development of diabetic retinal disease.

Disease	Manifestation							
	NVE	NVD	Artery Color	Vein Color	Artery Narrowing	Artery Dilation	Vein Narrowing	Vein Dilation
Choroidal Neovascularization	x							
Diabetic Retinopathy	x	x	x		x		x	x
Hypertensive Retinopathy				x	x			
Branch Retinal Vein Occlusion		x		x	x	x		x
Central Retinal Vein Occlusion		x		x	x	x		x
Hemi-Central Retinal Vein Occlusion				x	x			x
Branch Retinal Artery Occlusion		x	x		x		x	
Cilio-Retinal Artery Occlusion		x			x		x	
Arteriosclerotic Retinopathy			x					
Coats		x		x	x	x	x	x
Macro-aneurysm					x			
Hollenhorst Emboli					x			

Figure 1: Manifestations of diseases that affect the blood vessels of the retina (NVD = Neovascularization of the optic disk, NVE = Neovascularization elsewhere) [1]

The importance of focusing research on the study of retina is due to the fact that it is the only place where blood vessels and the tissue they perfuse can be directly and easily visualized.

There's a close relationship between retina and brain, such as to suggest that the health of the retinal vasculature reflects the vascular health of the brain and the retinal microvascular abnormalities are correlated with cerebrovascular diseases.

A large scale of studies have confirmed that retinal vascular abnormalities are generally associated with cardiovascular mortality.

Retinal photographs don't provide information just about diabetic retinal diseases but they also provide a significant amount of additional information relating to the entire vascular health of the individual. An ophthalmologist examines the vascular modifications by direct inspection or by manually analysing from the ocular fundus images; this inspection should be time consuming, expensive and prone to human error. Human measurements can produce in fact both imprecise results and problems concerned with the repeatability of the measurements.

Automated diagnosis of retinal fundal images using digital image analysis offers huge potential benefits. In a research setting, it offers the potential to examine a large number of images with time and costs savings and offer more objective measurements

than current observer-driven techniques.

Advantages in a clinical context include the potential to perform large numbers of automated screening for conditions such as diabetic retinopathy, and hence to reduce the workload required from manual trained graders [2].

Research goals and approach

In order to improve the accuracy of the measures that already exist, the *Computer Vision Team* within the School of Computing (University of Dundee) have developed retinal images analysis software since 2002 and have a suite of algorithms automatically taking measurements of various parameters, vascular characteristics, e.g. bifurcations, angles, thickness, tortuosity estimates and lesions, e.g. microaneurysm.

My main target was to develop a model that could well follow the intensity profile of a retinal vessel, and to measure the width of the vessel's cross section.

In literature many models have been developed with this purpose (the Gaussian Model, the Double Gaussian Model and the Hermite Model) but all of them present some inaccuracies and limitations. My goal is to obtain a model which fits the intensity profile, also when the profile is not symmetric or presents a central reflex, typical of some arteries.

I started studying the Multiresolution Hermite Model, which was developed by Li Wang to analyse retinal vasculature. Then, I modified it in order to overcome some of its limitations; obtaining an extended Multiresolution Hermite Model.

Once the model has been developed, I proceeded with tracking the intensity of retinal profiles, using the PSO (*Particle Swarm Optimization*) algorithm, which optimizes the model.

Subsequently the width of the cross section has been obtained implementing the FWHM (*Full Width Half Maximum*) technique.

The following step was the model validation. The statistics results demonstrated that the FWHM algorithm underestimated the width; then we improved the algorithm to measure the width. The new algorithm was called FW60M. Thus a new statistical analysis was required. It demonstrated that the model could returns widths comparable with the observers' ones.

Thesis outline

In **Chapter 1** some clinical aspects will be introduced. These are important to understand the rationale behind the focus of the study on retinal vessels.

In **Chapter 2** I will present some previous work completed by several authors, regarding the fit of intensity retinal vessel profiles, and the algorithms to measure the width of the retinal vessels.

In **Chapter 3** I will introduce the model that we decided to extend; I will explain step by step how we have achieved the Extended Multiresolution Hermite model. In the same chapter, I will introduce the optimization technique with the optimizer being called PSO.

In **Chapter 4** I will explain how we manually extract the intensity retinal profile; afterwards I will illustrate how the parameter's range values are chosen in order to cover a broad spectrum of intensity profiles. Finally I will explain how the PSO algorithm is been adapted to the model.

In **Chapter 5** I will describe two algorithms that were used from other authors to measure the vessel's width: the flex points technique and the FWHM algorithm. I will explain how I implement these algorithms in my work.

In **Chapter 6** I will discuss about model's validation. Some knowledge about the statistical methods that we utilized will be presented. Afterwards I will explain how we generate the ground truth, the protocols followed by the observers and the software to measure the widths. The statistical analysis will then be done and the outcome comes next.

In **Chapter 7** I will explain how I modified the FWHM algorithm in order to solve some problems that occurred during the validation phase. A new statistical analysis will follow.

In **Chapter 8** the results of my work will be presented, and finally some ideas for future analysis.

Chapter 1

Clinical background

1.1 Introduction

In this chapter some important clinical aspects will be introduced so that enable understanding to a greater extent the motivation behind the focus on retinal vessels.

1.2 The retina

The retina is the inner membrane of the eye; it consists of an intricate network of nerve-derived cells, starting with the rods and ending with the cones. Rods function mainly in dim light and provide black and white vision; cones support daytime vision and the perception of colour.

Rods and cones are photoreceptors, which detect characteristics of light entering the eye and transform it into electrical signals. After instantaneous processing within the retina, this visual information is sent on, and a picture is ultimately created in the brain.

The retina is only 1mm in thickness and it's the only portion of the body in which macroscopic blood vessels can be seen distinctly and evaluated. Because this crucial part of the eye is so metabolically active it's served by a second blood supply, the choroid.

In the centre of the retina there's the optic nerve, a circular to oval white area measuring about 2x1.5mm across. From the centre of the optic nerve radiate the major blood vessels of the retina.

Approximately 5mm to the left of the disc can be seen the slightly oval-shaped, blood vessel-free reddish spot, the fovea, which is in the center of the macula.

A circular field of approximately 6mm around the fovea is considered the central retina whereas beyond there is peripheral retina stretching to the ora serrata, 21mm from the

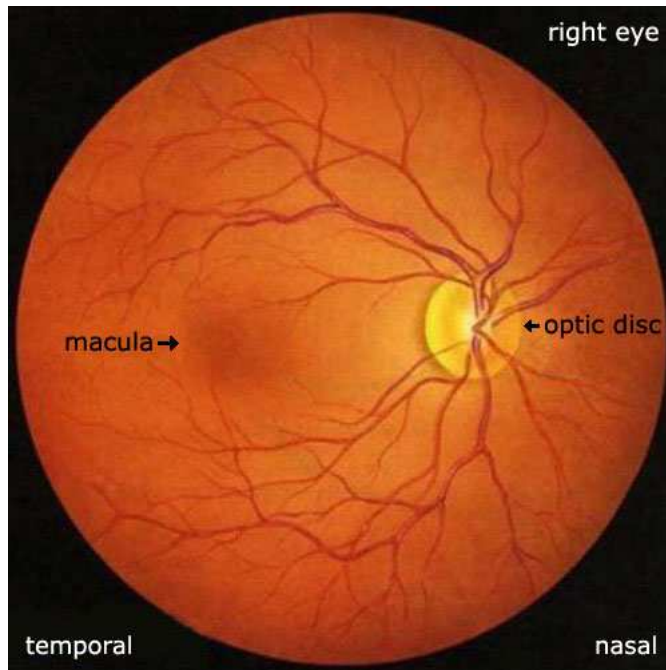


Figure 1.1: Retina

centre of the optic disc.

Overall, a typical retinal image from a fundus camera is a circular image covering a circular section in the eye of approximately 42 mm. This section is actually the projection of a concave surface.

The optic nerve contains the ganglionic cell axons running to the brain and, additionally, incoming blood vessels that open into the retina to vascularize the retina layers and neurons.

Light must travel through the thickness of the retina before striking and activating the rods and cones, subsequently the absorption of photons by the visual pigment of the photoreceptors is translated into a biochemical message and then an electrical message that can stimulate all the succeeding neurons of the retina.

The retinal message concerning the photic input and some preliminary organization of the visual image into several forms of sensation are transmitted to the brain from the spiking discharge of the ganglion cells.

1.2.1 Bloody supply to the retina

There are two sources of blood supply to the mammalian retina: the central retinal artery and the choroidal blood vessels.

The choroidal receives the greatest blood flow (65-85%) and it's vital for the maintenance of the outer retina and the remaining 20-30% flows to the retina through the central retinal artery from the optic nerve head to nourish the inner retinal layers.

The central retinal artery has 4 main branches in the human retina.

The arterial intraretinal branches then supply 3 layers of capillary networks:

1. The radial peripapillary capillaries
2. An inner layer of capillaries
3. An outer layer of capillaries

The precapillary venules drain into venules and through the corresponding venous system to the central retina vein.

The human retina is a delicate organization of neurons, glia and nourishing blood vessels; in some eye diseases the retina becomes damaged or compromised, and degenerative changes set in that eventually lead to a serious damage to the nerve cells that carry the vital messages about the visual image to the brain.

Diabetic retinopathy is a side effect of diabetes that affects the retina and can cause blindness. The vital nourishing blood vessel of the eye become compromised, distorted and multiply in uncontrollable ways [3].

1.3 Diabetes Mellitus

Diabetes mellitus, often referred to as diabetes, is a condition in which a person has a high blood sugar and glucose level, as a result of the body either not producing enough insulin, or because the body cells don't respond properly to the insulin that is produced.

Insulin is an hormone produced in the pancreas which enables body cells to absorb glucose and turn it into energy.

If the body cells don't absorb the glucose, the glucose then accumulates in the blood (hyperglycaemia) and this then can lead to various potential medical complications.

There are many types of diabetes, the most common are:

1. Type 1 Diabetes: results from the body's failure to produce insulin, thus requires the person to inject insulin.

It's characterized by loss of the insulin-producing Beta-cells of the isles of Langerhans in the pancreas leading to insulin deficiency.



Figure 1.2: A view of the fundus of the eye in a patient with advanced diabetic retinopathy

The majority of type 1 diabetes is of immune-mediated nature, where Beta-cell is a T-cell mediated autoimmune attack.

There's no preventive measure against type 1 diabetes, which causes approximately 10% of diabetes mellitus cases in North America and Europe.

Sensitivity and responsiveness to insulin are usually normal, especially in the early stages. Type 1 diabetes can affect children or adults but was termed "juvenile diabetes" because it represents the majority of the diabetes cases in children.

2. Type 2 Diabetes: results from insulin resistance, that is a condition in which cells fail to properly use insulin.

It's characterized by insulin resistance which may be combined with relatively reduced insulin secretion. The defective responsiveness of body tissues to insulin is improved insulin sensitivity or reduce glucose production by the liver.

Type 2 diabetes is the most common type.

In the early stage of type 2 diabetes, the predominant abnormality is reduced insulin sensitivity.

At this stage hyperglycemia can be reversed by a variety of measures and medications that improve insulin sensitivity or reduce glucose production by the liver. As

the disease progresses, the impairment of insulin secretion occurs, and therapeutic replacement of insulin may sometimes become necessary in certain patients.

Type 2 diabetes is determined primarily by lifestyle factors and genes: those who had high levels of physical activity, a healthy diet, didn't smoke and consumed alcohol in moderation had an 82% lower rate of diabetes. Obesity has been found to contribute to approximately 55% type 2 diabetes.

All forms of diabetes have been treatable since insulin became medically available in 1921, but treatment is difficult.

Pancreas transplants have been tried with limited success in type 1 diabetes mellitus (DM); gastric bypass surgery has been successful in many with type 2 DM.

Diabetes without proper treatments can cause many complications. Acute complications such hypoglycemia, diabetic ketoacidosis or nonketonic hyperosmolar coma. Serious long term complications include cardiovascular disease, chronic renal failure, retinal damage.

As of 2000, at least 171 million people worldwide suffer from diabetes, 2.8% of the population.

Type 2 DM is far the most common, affecting 90-95% of the United States diabetes population.

Diabetes mellitus is characterized by recurrent or persistent hyperglycemia and it's diagnosed by demonstrating anyone of the following:

- Fasting plasma glucose level at or above 7mmol/l
- Plasma glucose at or above 11.1mmol/l 2 hours after a 75g oral glucose load as in glucose tolerance test
- Symptoms of hyperglycemia and casual plasma glucose at or above 11.1mmol/l
- Glycated hemoglobin at or above 6.5

About a quarter of people with new type 1 diabetes have developed some degree of diabetic ketoacidosis by the time the diabetes is diagnosed.

The diagnosis of other types of diabetes is usually made in other ways, like ordinary health screening, detection of hyperglycemia during other type of medical investigations. Diabetes is often detected when a person suffers a problem that is frequently caused by diabetes, such as heart attack, stroke, neuropathy, poor wound healing or a foot ulcer, certain eye problems, certain fungal infections, or delivering a baby with macrosomia or hypoglycemia.

1.3.1 Epidemiology

In 2000, according to the World Health Organization, at least 171 million people suffer from diabetes (2.8% of the population).

Its incidence is increasing rapidly, and is estimated that by 2030 this number will be almost double; diabetes mellitus (especially type 2) is more common in the more developed countries. The increase in incidence of diabetes in developing countries follows the trend of urbanization and lifestyle changes.

For at least 20 years, diabetes mellitus in North America has been increasing substantially: in 2008 there were about 24 million people with diabetes in the United States only and other 57 million are estimated to have pre-diabetes.

The National Diabetes Information Clearinghouse estimated that diabetes costs \$132 billion in the US every year, about 5-10% of diabetes cases in North America are type 1, with the rest being type 2. The American Diabetes Association cites the 2003 assessment of the National Centre for Chronic Disease Prevention and Health Promotion that 1 in 3 Americans born after 2000 will develop diabetes in their lifetime.

1.4 Diabetic Retinopathy

Diabetic Retinopathy is a damage of the retina, caused by complications of diabetes mellitus (changes into the blood vessels of the retina), which can eventually lead to blindness.

It's an ocular manifestation of systemic disease which affect up to 80% of all patients who have had diabetes for 10 years or more.

Despite these intimidating statistics, research indicates that at least 90% of these new cases could be reduced if there was proper and vigilant treatment and monitoring of the eyes.

When blood vessels in the retina are damaged they may leak blood and grow fragile, brush like branches and scar tissue. This can blur or distort the vision images that the retina sends to the brain.

Proliferative retinopathy describes the changes that occur when new, abnormal blood vessels begin growing on the surface of the retina; this abnormal growth is called neovascularization. These new vessels have weaker walls and may be break and bleed, or cause scar tissue to grow that can pull the retina away from the back of the eye. When the retina is pulled away it is called retinal detachment and if left untreated, a retinal detachment can cause several vision loss, including blindness.

Leaking blood can cloud the vitreous and block the light passing through the pupil to the retina, causing blurred and distorted images.

In more advanced proliferative retinopathy, diabetic fibrous or scar tissue can form on the retina.

Diabetic retinopathy has four stages:

1. *Mild Non Proliferative Retinopathy*: at this earliest stage, microaneurysms occur. They are small areas of balloon-like swelling in the retina's tiny blood vessels.
2. *Moderate Non Proliferative Retinopathy*: some blood vessels that nourish the retina are blocked.
3. *Severe Non Proliferative Retinopathy*: many more blood vessels are blocked, depriving several areas of the retina with their blood supply. These areas of the retina send signals to the body to grow new blood vessels for nourishment.
4. *Proliferative Retinopathy*: the signal sent by the retina for nourishment triggers the growth of new blood vessels. These new vessels are abnormal and fragile, they grow along the retina and along the surface of the clear, vitreous gel that fills the inside of the eye.

By themselves, these blood vessels don't cause symptoms or vision loss. However they have thin walls, if they leak blood, severe vision loss can result [4].

Blood vessels damaged from diabetic retinopathy can cause vision loss in two ways:

1. Fragile blood vessels can develop and leak blood into the centre of the eye (blurring vision), this is proliferative retinopathy, as we just said.
2. Fluid can leak into the centre of the macula (the part of the retina responsible for the sharp, straight-ahead vision); the fluid makes the macula swell, blurring your vision, this condition is called macula edema; it can occur at any stage of diabetic retinopathy.

1.4.1 The Pathogenesis of Diabetic Retinopathy

Hyperglycemia appears to be a critical factor in the aetiology of diabetic retinopathy and initiates downstream events including basement membrane thickening, pericyte drop out and retinal capillary non-perfusion; of particular importance in the development and progression of diabetic retinopathy is the role of growth factors (e.g. vascular endothelial growth factor, placenta growth factor and pigment epithelium-derived factor) together

with specific receptors and obligate components of the signal transduction pathway needed to support them.

Diabetic retinopathy is characterized by hyperglycemia, basement membrane thickening, pericyte loss, microaneurysm, IRMA and preretinal neovascularisation which can eventually lead to blindness through haemorrhage and tractional retinal detachment[5].

Hyperglycemia is associated with a variety of biological events identified in the progression of diabetic retinopathy.

After 6 weeks that hyperglycemia has appeared, a proliferation of endothelial cells and swollen retinal vessels are observed; really interesting is that the retinal vessels lesions persist even after the blood glucose levels have turned into normal.

It's clear that there is a time point after which the progress of diabetic retinopathy is unavoidable and emphasizes that it is crucial to elicit preventive measures, such as intensive blood glucose control, at very early stages of diabetes, to prevent or slow progression of retinopathies.

In diabetes, early hyperglycemia is sufficient to increase the synthesis of basement membrane components in the retina which may contribute to the closure of capillaries. Diabetic basement membrane thickening appears to involve qualitative alternations of specific basement membrane markers at an advanced disease stage, with the appearance of diabetic retinopathy.

1.5 Retinal Vessel Structure

Reliable detection and segmentation of vasculature in retinal fundus images is important to different applications including image-guided therapy and automated oximetry.

A major challenge in vessel modelling is handling the variability in the vessel appearance [6].

The light emitted from a fundus camera is reflected back by the retina and choroids; blood vessels absorb some of the reflected light, absorbing into the wall and the blood. When an image of a blood vessel is viewed from point C (see Figure 1.3) the apparent width of the vessel is believed to be the width of the streaming column of erythrocytes viewed perpendicular to the flow direction (point P); the surrounding plasma zone and the vessel wall are transparent.

Due to the measured indices of refraction in the plasma and vitreous, it is accepted that within wide physiological limits the apparent width of the blood column is proportional to the true width.

Looking at the retinal blood vessels as a vascular cross-section, the overall trans-

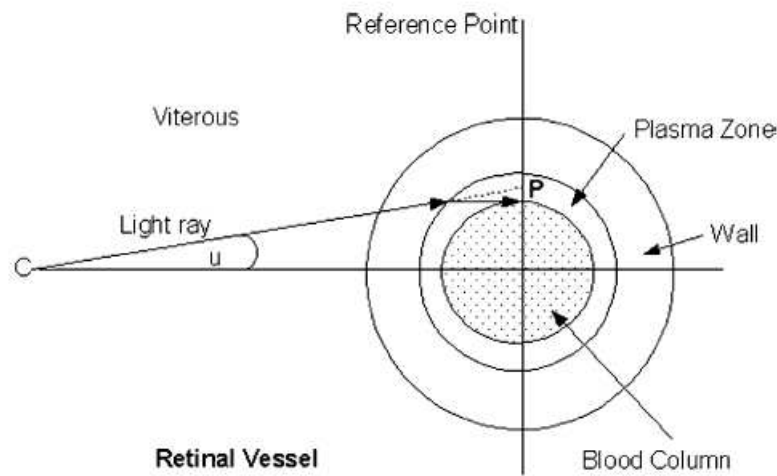


Figure 1.3: The apparent and true width of the blood column [7]

mittance through a vessel, including the blood column and wall, resembles a Gaussian curve.

Finding the point that the blood column meets the wall on that curve is extremely difficult, and to complicate the problem some vessels include a light streak known as light reflex that runs down the centre of the blood vessel; that is believed to be generated from a rough reflecting surface and the intravascular column of erythrocytes. Light reflexes are more common in younger retinas due to their increased reflective surfaces [7].

The phenomenon is pronounced at longer wavelengths which are more responsive to the blood oxygen content, particularly in arteries that are optically less dense than veins at these wavelengths.

In the artery, the reflex is not only stronger but appears to cover more of the vessel centre, the central reflex interferes with the measurements of the vessel optical density and must be avoided when carrying out manual and automatic analyses that incorporate density information, see Figure 1.4. The central reflex could create ambiguities so that the algorithm traces the two walls of the artery as separate vessels [6].

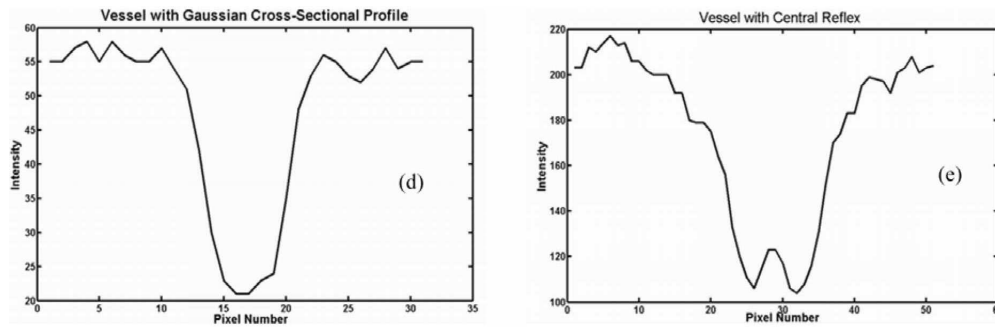


Figure 1.4: Vessel without central reflex and vessel with central reflex [6]

1.6 Conclusions

In the current chapter, I discussed some preliminary and research rationale clinical aspects. First, I presented some background about the retina and its vessels; next, some aspects regarding diabetes mellitus are presented, then an introduction about diabetic retinopathy and how this disease affects the retina and its vessels.

Finally, an explanation was provided about the typical retinal vessel structure, and the reason why the central light reflex, that occurs in some intensity profiles, is existing.

Chapter 2

Previous work for measuring retinal vessel profiles

2.1 Introduction

Vascular response to different physiological pathologies has been widely studied using a variety of methods. Measurements have been made both directly using ophthalmoscopy and indirectly using fundus photographs.

As previously mentioned, direct methods requiring ophthalmoscopic observations are often inaccurate due to non-standardized illumination, low magnification and a subjective bias.

More objective measurements are possible from fundus images in which the film transmission has been converted to intensity. Converting the film transmission to intensity has been described as the only objective measure of the properties of the photographed eye; however the observer has a restricted ability in discriminating between grey intensity levels [7].

In order to obtain objective measurements and reproducibility, several authors have presented algorithms for measuring vascular diameters. A presentation of these algorithms will follow.

2.2 Introduction to diameter measurement algorithms

Direct methods requiring ophthalmoscopic observations are often inaccurate due to non-standardized illumination, low magnification and a subjective bias. More objective measurements are possible from fundus images in which the film transmission has been converted to intensity; this one is the only objective measure of the properties of the photographed eye.

A vascular cross-section is often presented in graphical form as Figure 2.1.

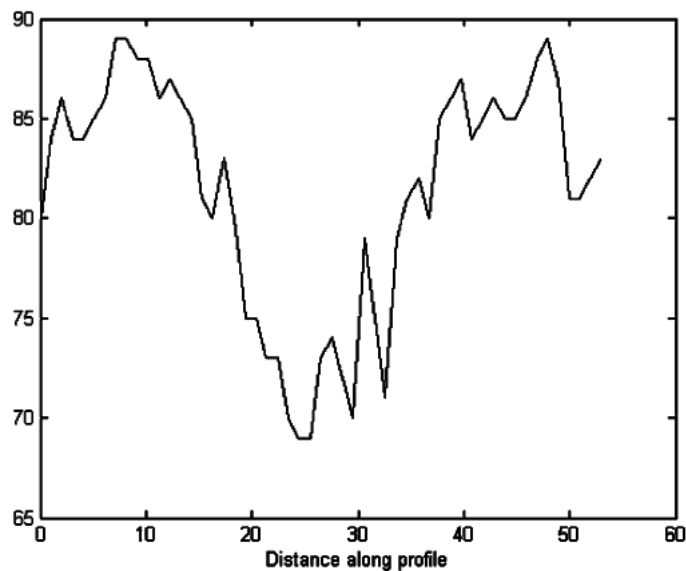


Figure 2.1: Sample vessel profile [7]

The intensity level of the back-ground retina is often different on either side of the vessel and the vessel often contains a bright reflex.

For automated diagnostic purposes, obtaining consistent measurements is more important than obtaining the actual vessel widths, as it deals with variations in widths along a vessel, or ratios of widths at difference points in the vessel of interest, rather than the absolute value of width [7].

Several models have been presented for fitting retinal vessel profile:

- *Simple Gaussian*
- *Double Gaussian*
- *Hermite Model*

and algorithms have been presented by previous authors for measuring vascular diameters:

- *Brinchmann-Hansen*, they presented an algorithm called full-width half-maximum (*FWHM*) to measure vascular diameters
- *Gregson's* approach comprising of a rectangular profile of fixed height that is fitted to the profile data

2.3 Models for fitting retinal vessel profiles

2.3.1 1-D Gaussian Model: Zhou and al.

Liang Zhou presented an algorithm for the analysis and quantification of the vascular structures of the human retina.

The algorithm relies on a matched filtering approach coupled with *a priori* knowledge about retinal vessel properties to automatically detect the vessel boundaries, track the midline of the vessel and extract useful parameters of clinical interest; that is obtained using Gaussian functions [8].

Some interesting properties of blood vessels in the retinal images have been observed:

1. The vessel direction varies continuously. The vessel consists in an unlimited number of segments and corresponding direction vectors with no discontinuity between two adjacent segments.
The change of direction between segments is a smooth continuous function.
2. The vessel width varies continuously; there is always a smooth transition between adjacent segments.
3. The density distribution of a blood vessel cross-sectional profile can be estimated using a Gaussian shaped function.

Then a Simple Gaussian fitting function has been chosen to model the filter curve because it fits the density distribution in most situations.

The matched filter helps to eliminate errors due to small branches at a bifurcation point, thus allowing tracing to follow one major branch continuously.

A representative grey level distribution of retinal artery profile is illustrated in Figure 2.2

The vessel profile is been modeled using a Gaussian function of the formula:

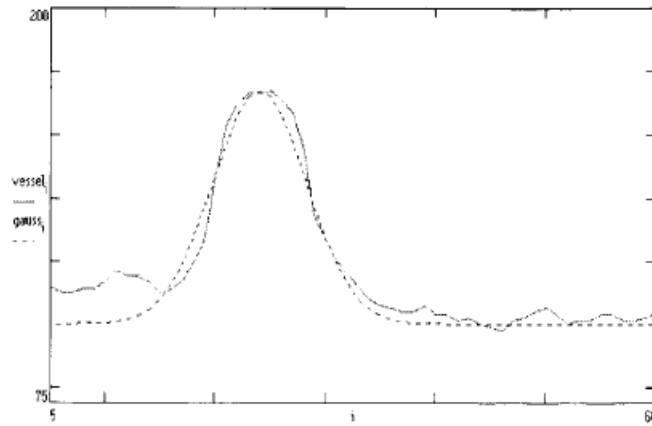


Figure 2.2: Typical vessel density profile with no bifurcation and its matched filter curve

$$g(l) = Ae^{\frac{-(l-c)^2}{2\sigma^2}}$$

where:

- l is the horizontal distance between two points $(l, g(l))$ and the origin point
- c is the location of the center of the Gaussian function
- σ is the standard deviation, it defines the spread of the intensity profile
- A represents the maximum grey-level intensity near the center of the vessel

The determination of the vessel edge is difficult; both tails of the Gaussian function are cut at the 95% point, which means that pixels outside the two tails are not considered as vessel pixels, so the matched filter curve is given by:

$$f(l) = \begin{cases} Ae^{\frac{-(l-c)^2}{2\sigma^2}} & \text{if } -1.96 < \sigma < 1.96 \\ 0 & \text{otherwise} \end{cases}$$

This type-matched filter has the following advantages [8]:

1. it has a reproducible definition for the edge of a vessel
2. at bifurcations nodes the matched filter discriminates against the weaker signal
3. Gaussian matched filter gives better accuracy in locating the edge, compared to Gate's filter [8].

2.3.2 2-D Gaussian Model: Lowell, Hunter et al.

Lowell, Hunter et al. presented an algorithm to measure the vessel diameter to subpixel accuracy; their algorithm is based on a two-dimensional model with a Gaussian or Difference of Gaussian profile, which is iteratively optimized to best fit the observed vessel. This allows to calculate the vessel width to subpixel accuracy and the smoothing introduced by the two-dimensional nature of the profile improves this accuracy.

The model is fitted on an intensity image produced by extracting the green channel from an original color digital image; the method may be integrated with any vessel segmentation algorithm and with any algorithm to diagnose pathology on the basis of vessel calibre.

The 2-D model consists of an idealized straight vessel segment, with a Gaussian profile, G1, orthogonal to the vessel direction. The size and shape of the model is controlled by a small number of parameters, which can be adjusted to fit the model to the actual vessel profile; it is extremely important that the model mimics the vessel segment being measured as accurately as possible; a variation is used to model the light reflex that could be found in some vessel segments.

The 2-D model without light reflex (see Figure 2.3) has a profile shape governed by Equation 2.1 and Equation 2.2

$$\alpha = x \sin \theta - y \cos \theta - \mu \quad (2.1)$$

$$f(x, y) = t - h_1 e^{-s_1 \alpha^2} \quad (2.2)$$

where:

x, y map the profile maximum

t is the profile maximum

h_1 is the height of the Gaussian

s_1 is the width of the Gaussian ($= 1/\sigma_1^2$)

μ is the offset of the Gaussian center from the initial center estimate, orthogonal to the orientation.

The light reflex is achieved by subtracting a second small Gaussian curve G2 from the main one; the second Gaussian is oriented as the same angle as the first and is independently adjustable allowing the overall model to fit well to both blood vessel column wall and light reflex then the 2-D model with light reflex shares the same base parameters as its non light reflex counterpart; additional parameters are included to control the shape and size of the light reflex.

The model with light reflex (see Figure 2.4) is expressed by the Equation 2.3:

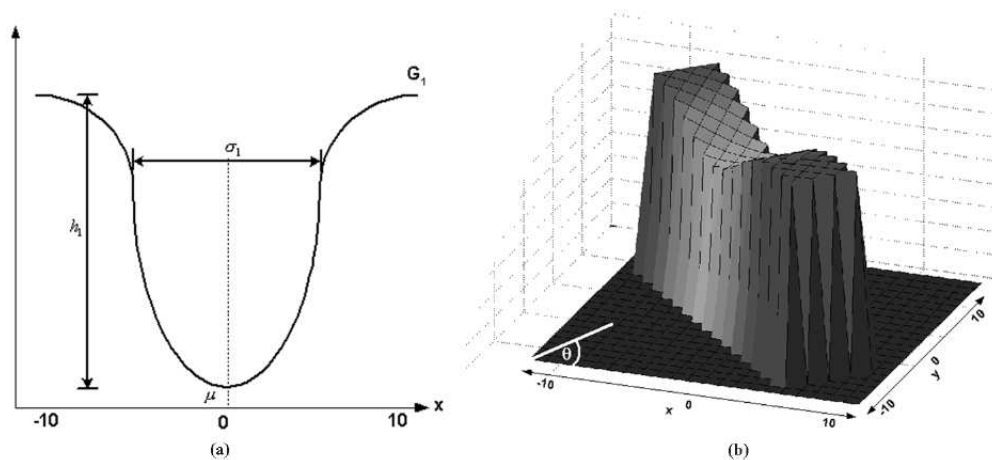


Figure 2.3: Model without light reflex a) Cross-section profile b) Two-dimensional representation [7]

$$f_{x,y} = t - (h_1 e^{-s_1 \alpha^2} - h_2 e^{-s_2 \alpha^2}) \quad (2.3)$$

with h_2 and s_2 set the height and the width of the Gaussian G_2 respectively.

It has been verified that the difference in standard deviations between the 2-D model and the alternative approaches is statistically significant and, using F-test, that the 2-D method is better than the average 1-D model.

Zhou and al. claimed that their approach, the 1-D Gaussian model was more precise than other algorithms; however the 2-D algorithm is 32% more precise. The 1-D model is also prone to failure to converge sufficiently to obtain a sensible width readings.

The use of a 2-D model circumvents problems caused by noisy data such as small branches and introduces a degree of smoothing that helps to improve the model fit, the inclusion of a Difference-of-Gaussian model improves also performance over a single Gaussian where there is a visible light reflex [7].

2.3.3 Multiresolution Hermite Model (MHM): Wang et al.

Wang et al. (2007) presented a vascular representation and segmentation algorithm based on a Multiresolution Hermite Model (MHM) for vascular segmentation and target it to the accurate segmentation of retinal vasculature.

A multiresolution model is a model which captures a wide range of levels of detail of an object, and which can be used to reconstruct any one of these levels on demand.

In ophthalmic practice, angiography is a common procedure and several different

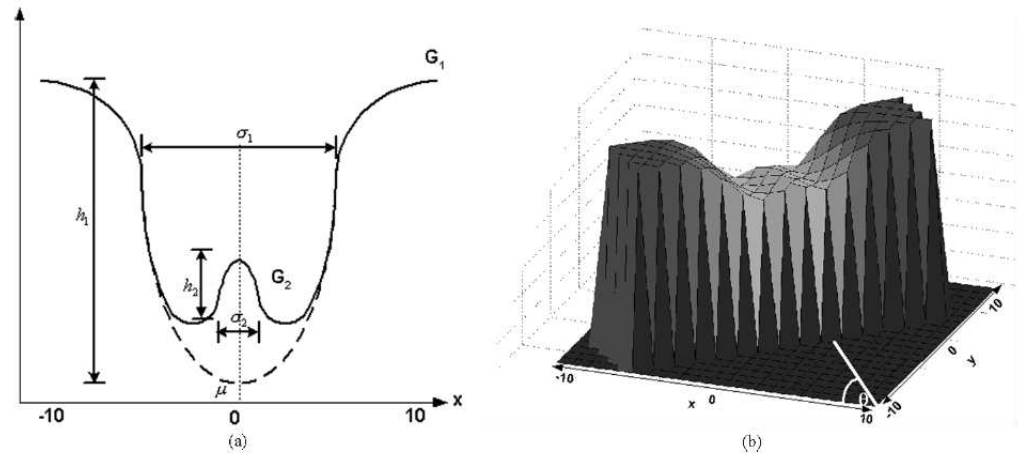


Figure 2.4: Model with light reflex a) One-dimensional representation b) Two-dimensional representation [7]

techniques are commonly used, such as fluorescein and indocyanine green angiography and digital fundus photography; this is used (without contrast agents) for screening hypertension and diabetes.

Nevertheless, the resulting images can pose significant challenges for human experts to interpret and, in the case of screening, manual assessment can be tedious and error prone. The hope is, therefore, that quantitative computer-based assessment of the characteristics and functionality of the blood vessels for these angiographies may be faster and yield more consistent results than a manual examination process.

Central is the detection of blood vessels in an image; this allows for a quantitative measurement of the geometrical changes of arteries, in their diameters, tortuosity, or lengths and can provide the localization of land mark points (bifurcations) needed for image registration.

Retinal vasculature has several distinct characteristics that can be used by image processing techniques to separate blood vessels of interests from the background.

First, vessel cross-sectional intensity profiles approximate a Gaussian in shape for fundus images. Second, the orientation and grey level of a vessel does not change abruptly; they are locally linear and gradually change in intensity along their lengths. Finally vessels can be expected to be connected and, in the retina, form a binary treelike structure [9]. However, the shape size and local grey level of blood vessels can vary hugely and some background features may have similar attributes to vessels.

Many segmentation algorithms have been presented to provide automated or semi-automated detection of vascular structure; these methods attempt to exploit one or

more of the characteristics of retinal vessel and can be categorized into: tracking-based approaches, neural network approaches, and template and model based approaches.

Tracking-based approaches are essentially region-growers, albeit designed specifically for linear structures; they require user intervention to determine the thresholds for vessel versus background, the region seeds point and the choice of thresholds during the tracing process.

Neither tracking and neural network classifiers extracts the vascular structure directly and a postprocess of line-thinning or skeletonization may have to be applied to estimate the global connectivity.

Line-thinning is problematic: as small variations in the local geometry in relation to the pixel sampling grid can lead to different estimates of the vessel's medial axis; thinning is also sensitive to noise [9].

The final results will lead to inaccuracies in width, vessel direction and tortuosity measurements.

Template and model based approaches used a set of two-dimensional Gaussian kernels with fixed size and an orientation used to enhance the vessels.

The method has some interesting characteristics, such as detecting bifurcations, to some extent and obtaining robust separation of vessels from background. However, it is relatively time consuming as the convolution kernel used may be quite large and needs to be applied repeatedly.

In addition vessels with high tortuosity may also be missed out in the kernel response; and finally a postprocess is still required to join up the filter responses either by tracking/region-growing followed by a skeletonization or by other labeling processes.

The Gaussian profile assumption can not deal with the central light reflex artifact, to address this problem several authors have suggested the extension of the profile model to use mixtures, Laplacian of Gaussian or Difference of Gaussian. This complicates the matched filtering since it conflates the problem of what size of Gaussian to use with the need to estimate a mixing and a second size parameter.

The solution of Wang et al. is to use a Hermite model that requires a single mixing parameter and generalizes their vessel profile model without the need to change the estimation strategy.

A two-dimensional Hermite function intensity model is developed which models blood vessel profiles in a quad-tree structure over a range of spatial resolutions.

The use of multiresolution representation simplifies the image modeling and allows for a robust analysis by combining information across scales. Estimation over scale also reduces the overall computational complexity.

Unique to their approach is the use of a topological prior that seeks out binary trees

in the image, this allows to extract a meaningful global description of the vasculature that is robust to noise and obviates the need for a postlabeling skeletonization step.

Wang's main goal had been to develop a branching structure segmentation and detection; to achieve this goal they have developed a vessel modeling and estimation technique that operates directly on the image intensities for detection of the blood vessels [9].

To account for changes in scale of the features of interest, a block-based multiresolution approach is employed, with an expectation-maximization (EM) type of optimization scheme to fit local model parameters. The local models are combined across scales to produce a data summary by a model selection. Then a stochastic Bayesian approach is used to link the local models and infer the global vascular structure under the assumption of a binary tree structure.

Blood vessel segments are approximately Gaussian in shape, however, in some special cases, the intensity profile of some large vessel segments is not Gaussian due to the *central light reflex*; this happens more frequently in digital fundus images, since the light reflex is comparatively larger than in fluorescein images. For the above special cases, Gaussian intensity modeling tends to over-segment the features.

The difference of two separate Gaussian function that had been proposed by Gao et al., is able to fit the data well, but doubles computational complexity.

Wang et al. proposed for solve this problem an approach that uses Hermite approximation which introduces only one additional parameter.

The Hermite function is based on the natural extension of Gaussian function, it is the orthogonalisation of the derivatives of the Gaussian.

The one dimensional (1-D) Hermite polynomial is defined as [9]:

$$H_n(x) = e^{x^2/2} \left(x - \frac{d}{dx} \right)^n e^{-x^2/2}$$

By Cartesian separability, its 2D extension can be represented as:

$$H_{mn}(\theta, a) = H_n(x)H_m(y)$$

The Hermite model consists of a second-order Hermite polynomial $H_{2,0}(x)$ used with an adoptive parameter "a" as a factor that is multiplied by the optimized Gaussian $G(x | \theta)$:

$$H_{2,0} = (1 + a(x^2 - 1))G$$

which a = scalar parameter perpendicular to the orientation of the Gaussian.

In Figure 2.5 we can see the 1-D plot of $H_{2,0}$, with different values of "a", when $a = 0$, $H \equiv G()$ and the two peaks are further apart when "a" increases.

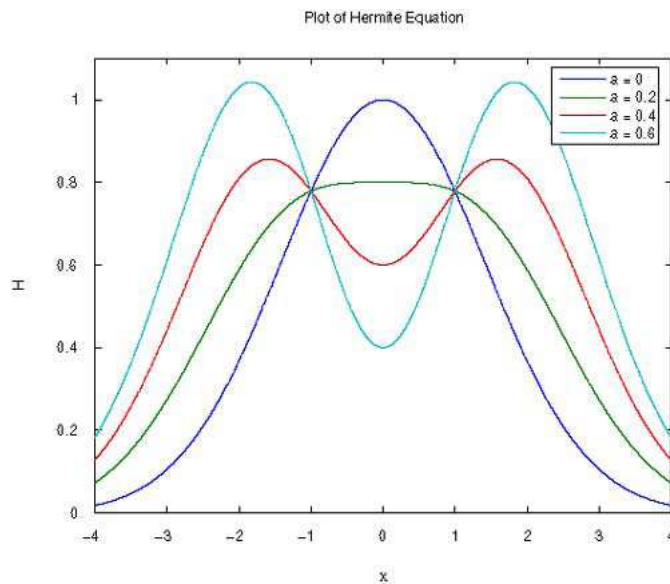


Figure 2.5: Plot of 1-D Hermite function with different a [9]

In Figure 2.6 it is possible to observe the Hermite model gives a more accurate approximation than a simple Gaussian.

The principle of model selection is to use the smallest possible number of parameters for adequate representation of the data [10]. The principle of model selection is a bias versus variance tradeoff. In general, bias decreases and variance increases as the number of parameters increases: the fit of any model can be improved by increasing the number of parameters, but a tradeoff with increasing variance must be considered in selecting a model.

Wang et al. use a hierarchical quad-tree structure to perform feature estimation at different levels, the Akaike Information Criteria (AIC) is used to bias the likelihood function with the number of parameters.

The MHM model has been shown to be effective and robust on vessel segmentation from noisy data; the vessel classification results using the MHM model were compared on two publicly available databases, STARE and DRIVE, and were shown to produce accurate results, outperforming the best available methods, especially in higher sensitivity rates [9].

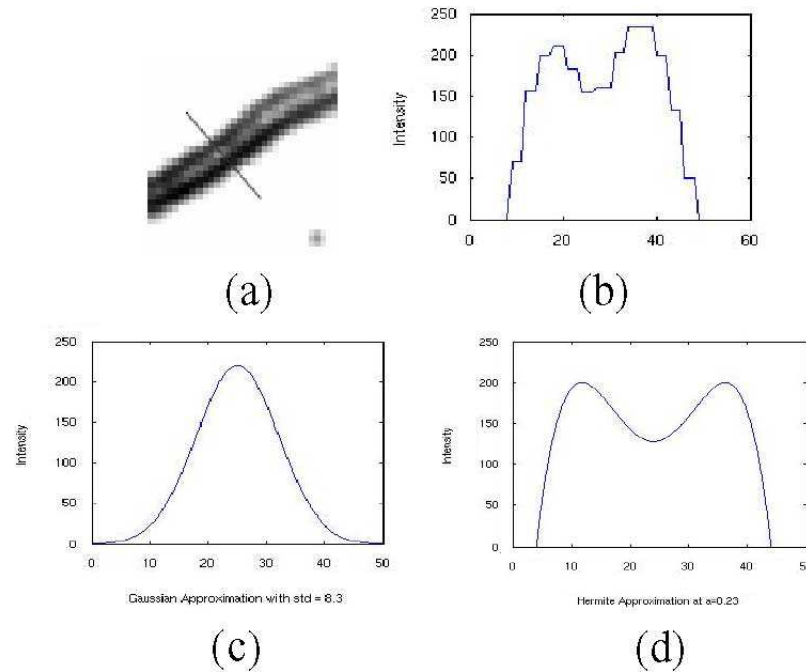


Figure 2.6: (a) Vessel section after background subtraction.(b) Plot of a cross section of the vessel intensity profile.(c)Plot of a Gaussian approximation.(d) Plot of a Hermite approximation [9]

2.4 Algorithms for measuring vascular diameters

2.4.1 FWHM (*F*ull *W*idth *H*alf *M*aximum): Brinchmann, Hansen et al.

Brinchmann and Hansen (1986) have measured the width of the blood column in retinal vessel using microdensitometry and found that changes in the width of the blood column may be related to pathological changes in the vessel wall; they investigated also the importance of the central light reflex on retinal vessels [11].

They used the full width half maximum method to determine the width of a vessel.

The width measurement has been based on the intensity profiles, orthogonal to a smoothed trace, in the vessel found by Dijkstra's shortest path algorithm (see Figure 2.7).

The intensity profiles show large varieties in shape. Often the background levels are different and an axial reflex (AR) is often seen in the centre of the profile (see Figure 2.8).

The success of the method is highly dependent on robust estimates of the back-

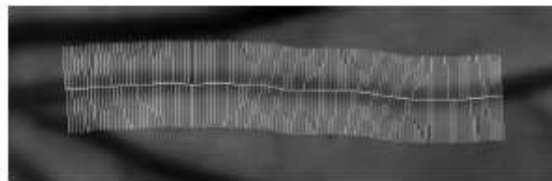


Figure 2.7: Profile lines, overlaid vessel segment. [11]

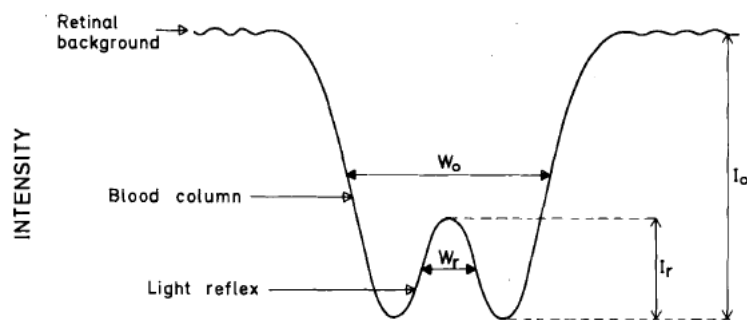


Figure 2.8: A schematic illustration of the various scan parameters in a densitometric profile across a retinal vessel. I_o , is the depth of the vessel absorption (intensity) relative to the average retinal background, and I_r , is the intensity of the light reflex. The width of the blood column W_o , and the width of the light streak W_r [12].

ground intensity levels.

The typical intensity profile orthogonal to a retinal vessel is shaped similar to a Gaussian curve; looking at Figure 2.9 there are two distinct minima levels (*MinLeftLevel* and *MinRightLevel*) due to the presence of an AR. The half maximum in the right side is calculated as the mean of the left background level (*MaxLeftLevel*) and the left minimum level (*MinLeftLevel*); the half maximum in the right side is vice versa.

The FWHM estimate of the profile width, \hat{D}_{FWHM} , is the full width between the two half maximum; this method is not sensitive to the presence of the AR.

$$\text{half minimum left} = (\text{max left} + \text{min left})/2$$

$$\text{half minimum right} = (\text{max right} + \text{min right})/2$$

\hat{D}_{FWHM} = is the full width between the two half minima.

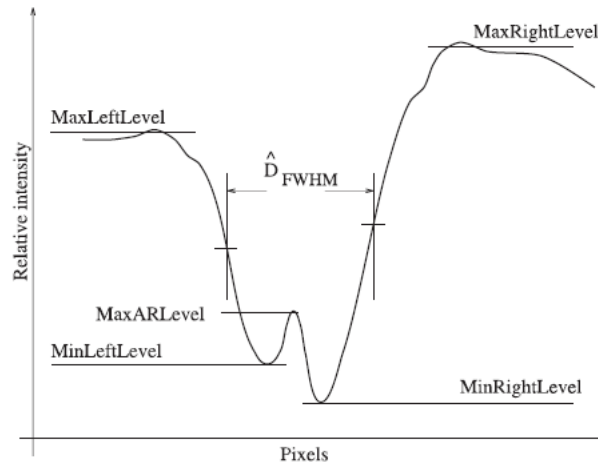


Figure 2.9: Principles of full width half maximum estimation vessel width The estimated profile width is the distance between the two vertical dashed lines. [11]

2.4.2 Kick points: Rassam et al.

Rassam et al. observed (1994) the phenomenon of two “kick points” on the slopes of the transmittance and densitometry profile.

They demonstrated that the kick points being coincident with the edges of the blood column; the horizontal distance across them indicate the actual blood column width [13].

To evaluate this hypothesis, blood was infused through two lengths of plastic tubing of known diameters, the half height underestimated the blood column width by 7.33% and 6.46%, while the kick point method slightly overestimated it by 1.40% and 0.34%.

These techniques were then applied to monochromatic fundus photographs, the kick point method may provide the most accurate measurement of vessel width.

The most commonly used techniques in determining vessel width from retinal photographs were microdensitometry, which measured the density profile, film scanners that recorded photographic image transmittance, and film densities to provide film exposure profiles.

The only available method to derive an estimate of the vessel width from these profiles, used the arbitrary assumption of the width at half height of the profiles as a measure

of the blood column width. This technique is dependent on the characteristic photosensitivity calibration curves of the photographic film used, this is a potential source of error.

Viewing the Figure 2.10 it is possible to observe two distinct skew points ("kick points") on the slopes in the transmittance profile near the base of the curves, these points are also seen in densitometry profiles [13].

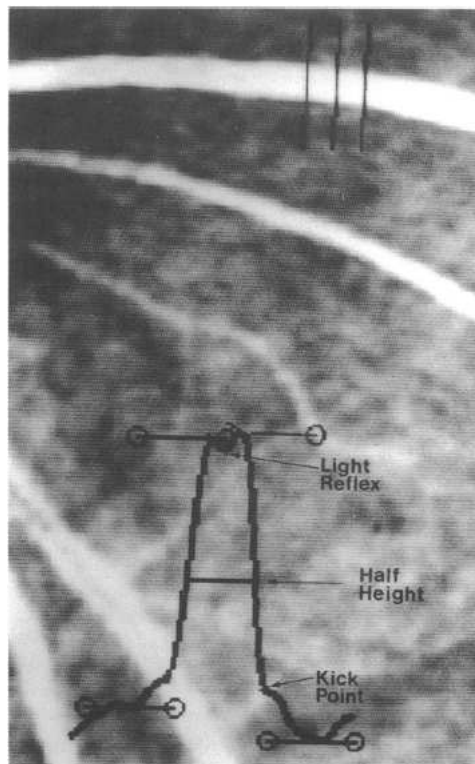


Figure 2.10: Monochromatic fundus photograph, displaying the transmittance profile of a vessel from the photographic negative. The half-height, the kick points and the light reflex are shown [13]

These points mark the edges column; then taking the horizontal distance from one kick to the other, an accurate measurement of the width of the blood column can be obtained.

The kick points are dependent on two facts:

- in the light path through the vessel wall is greatest at the lateral interface between the blood column and the vessel wall
- the absorption coefficient is different for the blood from that of the vessel wall.

If the blood and vessel wall have the same optical densities then the kick points will not appear, this is the main problem of this technique. To observe the kick points, the photograph must be high resolution and well confused, as the shape of the intensity curve changes with focusing.

This makes the detection of the kick points difficult [7].

2.4.3 SLRF method (Sliding Linear Regression Filter): Chapman et al.

Chapman et al. (2001) have compared three automated methods and a manual method for measurement of arteriolar diameters from digitised red-free retinal photographs.

Some diameters, measured by manual identification of vessel edges in pixels, were compared with diameters measured by:

- fitting vessel intensity profiles to a double Gaussian function by non-linear regression. The intensity cross-section of the vessel was modelled by

$$I(x) = a_1 e^{-\left(\frac{x-a_2}{a_3}\right)^2} + a_4 - a_5 e^{-\left(\frac{x-a_2}{a_7}\right)^2}$$

- a standard edge detection algorithm (Sobel)
- and determination of points of maximum intensity variation by a sliding linear regression filter (SLRF)

The SLRF method is based upon the fitting of a line by linear regression, relating image intensity against distance along the cross-section, within a window W points centred on the n th point.

The window is progressively moved by a single point at a time across the entire section of interest, the width of W is determined by a moment calculation; at each position of the window the gradient of intensity change with distance is measured by a linear regression calculation, then the SLRF of the best least squares line recorded against the n th point as m_n , see Figure 2.11.

To determine the positions of the maximum positive going and negative going slope, the resulting values of m are subject to a threshold test, and those points falling beyond a specified range are used to compute the actual position of the edge, see Equation 2.4:

$$x_{edge} = \frac{\sum_i x_i m_i}{\sum_i m_i} \quad (2.4)$$

for every i : $m_i > k_i m_{max}$ and $m_i < k_i m_{min}$

where:

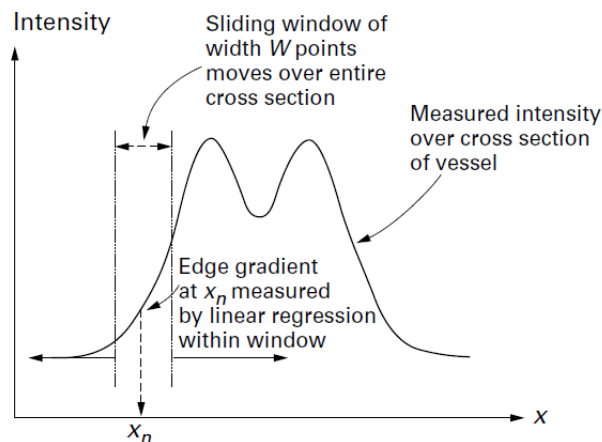


Figure 2.11: The sliding linear regression filter (SLRF) is applied to a vessel cross section[14]

m_{max} = maximum value of the slope

m_{min} = minimum value of the slope

k_i = a constant (~ 0.5)

The vessel width is given by the difference between x_{edge} values.

A further consideration in the SLRF method is the choice of window size W : a smaller window size will give better resolution of edges of small vessels, but at the expense of less noisy rejection; the window size should be adapted to reflect the size of the vessel under examination.

Concluding the SLRF method appears to be the most consistent and repeatable, but it requires a sufficient number of points within each window to calculate a reliable estimate of slope by linear regression; the method becomes unreliable for total vessel widths of less than 10 pixels.

2.5 Conclusions

The present chapter is divided into two parts.

In the first part, I presented some algorithms that were used in previous studies to fit the intensity vessel profile: a simple Gaussian Model, a 2-D Gaussian Model and a Multiresolution Hermite Model. All these algorithms showed to present some inaccuracies and limitations: they can fit only symmetric profiles and they have some limitations

to fit the central light reflex. Due to these reasons, we decided to extend one of these algorithms in order to improve it.

In the second part, some algorithms used for measuring vascular diameters were presented: Full Width Half Maximum (FWHM), kick points and the Sliding Linear Regression Filter (SLRF) method.

Chapter 3

Extension of the Multiresolution Hermite Model

3.1 Introduction

In this chapter, I explain how the Extend Multiresolution Hermite Model has been obtained. Starting from the Wang's model, some parameters with specific functions were added. In the last part, an introduction about PSO algorithm was added, the PSO is used to find the best value for each parameter.

3.2 Why we decided to extend the MHM

The MHM (MultiResolution Hermite Model) that had been used by Wang :

$$H_{2,0} = [1 + a(x^2 - 1)]G$$

showed to fit very well with the intensity profile of retinal vessels, due to its ability to represent the possible central reflex of some vessels.

The results of Wang's process demonstrated that the MHM is very efficient in terms of sensitivity, it has high level of agreement with the observers on which pixels are part of the vasculature.

Adopting a multiresolution approach is a key aspect in global image analysis, this means that the local region analysis has to be applied to regions of varying sizes to find the optimal scale to model particular features. The use of multiresolution representation simplifies the image modeling and allows for a robust analysis by combining information across scales; estimation over scale also reduces the overall computational complexity

[9].

Due to its good properties we decided to extend the MultiResolution Hermite model.

3.3 Experimental model setup

We started with a deep analysis of the MHM formula suggested by Wang:

$$H_{2,0} = [1 + a(x^2 - 1)] \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

where, as explained in previous chapter:

- a is a scalar parameter perpendicular to the orientation of the Gaussian when 2-D modeling has been performed, $a \in [0, 0.8]$
- m is the mean of the Gaussian
- σ^2 is the variance of the Gaussian

The formula is composed of two parts: an Hermite polynomial and a Gaussian function.

The polynomial

$$[1 + a(x^2 - 1)] \tag{3.1}$$

which has a graphical representation, as shown in Figure 3.1:

The Gaussian function

$$G = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{3.2}$$

which is represented with mean $m = 0$ and $\sigma^2 = 1$ in Figure 3.2:

If we combine these two functions 3.1 and 3.2 we obtain the Hermite function

$$H(a, m, \sigma, x) = [1 + a(x^2 - 1)] \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{3.3}$$

and Figure 3.3 represents the Hermite function:

Looking at Figure 3.3, we can see that, due to the different value of parameter a , different trends can be obtained; they are different for the greatness of the central deformation.

The first restriction shown by these trends is that they have axial symmetry with the axis passing by their mean value m ; this constraint implies that all the intensity vessel profiles are symmetric with their centerline, yet this is not always the case.

Thus we introduced a new parameter δ in our polynomial, then a formula has been developed in Equation 3.4:

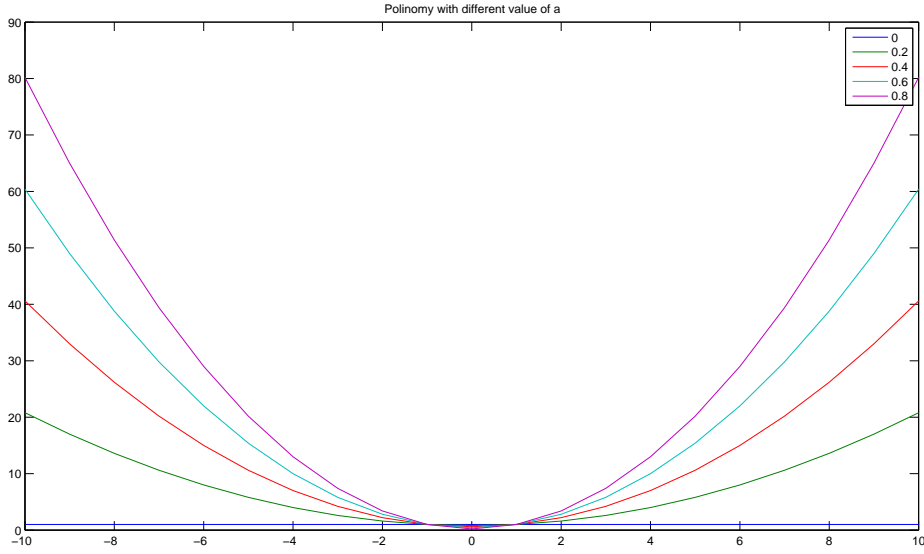


Figure 3.1: Trend of the Hermite polynomial

$$H(a, m, \delta, \sigma, x) = \{1 + a[(x - m + \delta)^2 - 1]\} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (3.4)$$

This formula allows to obtain different trends, due to a , but also trends with different symmetries, due to δ , see Figure 3.4.

Another multiplier parameter p has been added to this formula, this parameter involves also the term $\frac{1}{\sqrt{2\pi\sigma^2}}$. Due to the change in sign of the function, a negative value for p has been chosen, the modified formula can be seen in Equation 3.5 and its trends in Figure 3.5.

$$H(a, m, \delta, \sigma, p, x) = p\{1 + a[(x - m + \delta)^2 - 1]\} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (3.5)$$

The values in the y-axis can not be negative because they represent the intensity values of the pixels of the cross section under analysis, that meaning, that they include values from 0 to 255.

Due to the reason that parameter q (the last parameter) can shift the function along the y-axis, it has been added to the formula.

Therefore the final expression for the Extended Multiresolution Hermite Model is:

$$H(a, m, \delta, \sigma, p, q, x) = p\{1 + a[(x - m + \delta)^2 - 1]\} e^{-\frac{(x-m)^2}{2\sigma^2}} + q \quad (3.6)$$

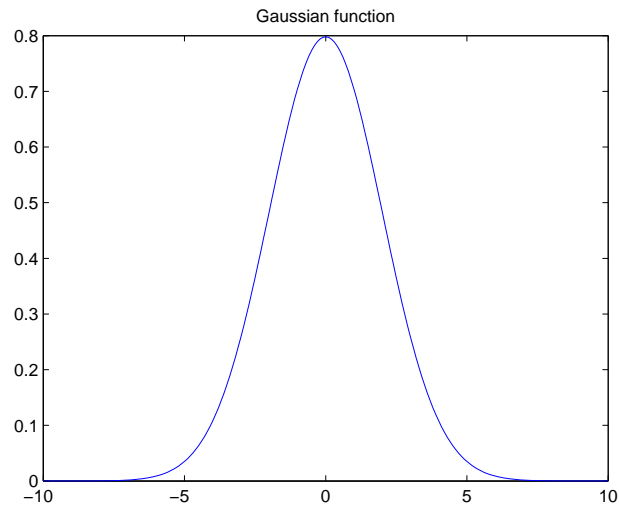


Figure 3.2: Trend of the Gaussian function

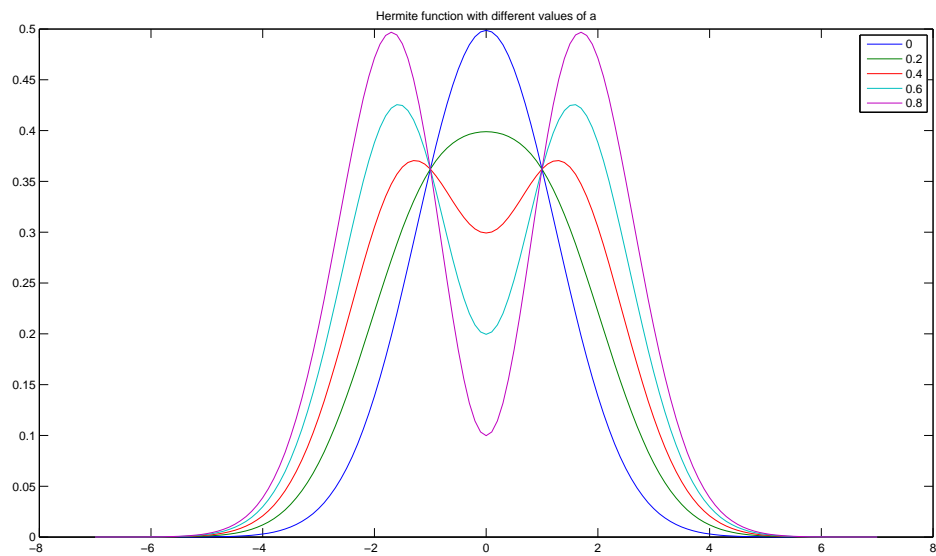


Figure 3.3: Trends of the Hermite function

For example, if we assign the following values to the parameters: $m=0$, $\delta=0.2$, $\sigma=1$, $p=-150$ and $q=255$, we obtain a fit like the one shown in Figure 3.6.

We are interested in obtaining a generical model, that considers all the possible intensity retinal vessel profiles such as the simple gaussian profiles, profiles with central

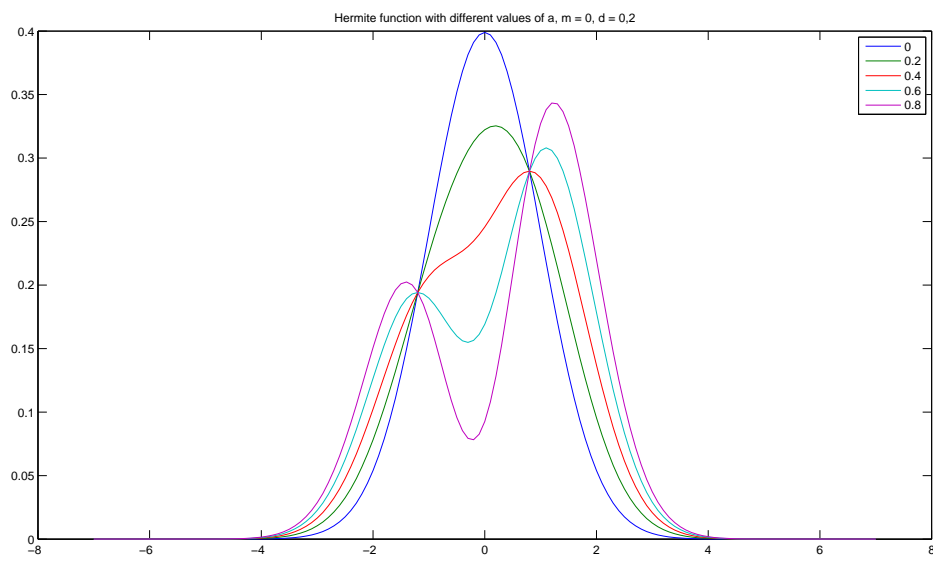


Figure 3.4: Trends of Equation 3.4

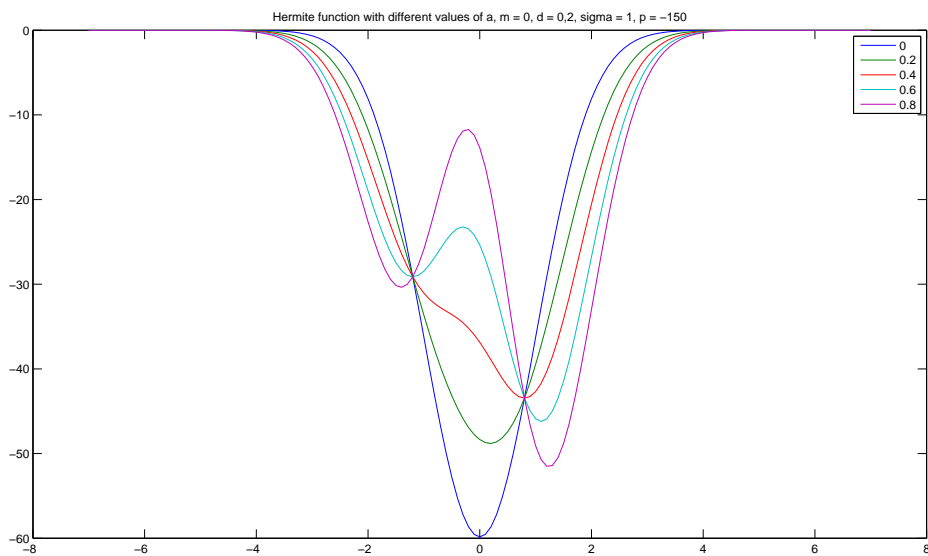


Figure 3.5: Trends of Equation 3.5

reflex, symmetrical and asymmetrical profiles.

Through intensity profiles based observations, we have estimated the optimum range of values for each parameter:

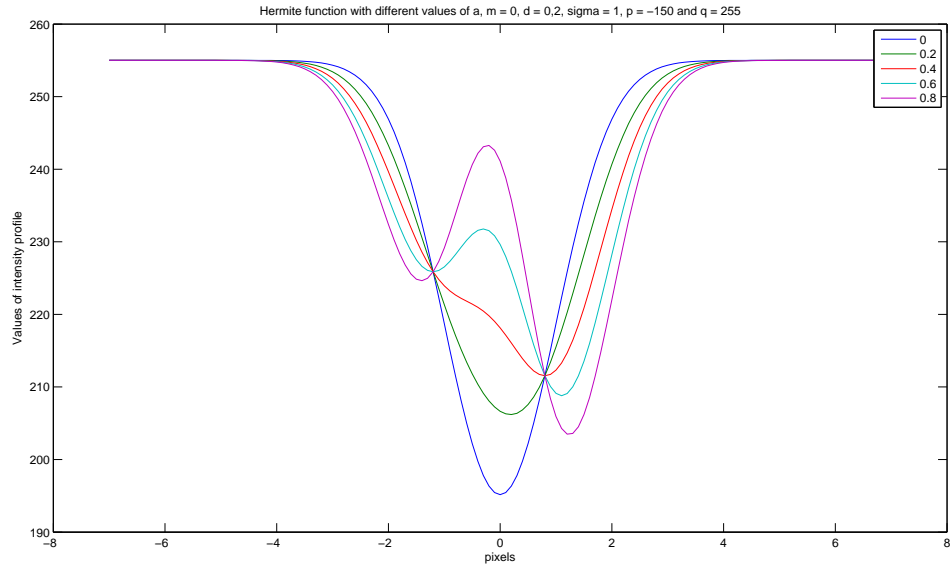


Figure 3.6: Trends of Equation 3.6 with the assigned values: $m=0$, $\delta=0.2$, $\sigma=1$, $p=-150$ and $q=255$

$$\begin{aligned}
 a &= [0 \ 0.8] \\
 m &= [7 \ 50] \\
 \delta &= [0.01 \ 2] \\
 \sigma &= [1 \ 23] \\
 p &= [-150 \ 0] \\
 q &= [255 \ 0]
 \end{aligned}$$

Taking these ranges, we can cover a broad spectrum of intensity retinal vessel profiles.

In the next chapter, I will discuss the role of this parameters in the global trend of the Extended Multiresolutional Hermite Model.

3.3.1 Summary of the Extended Hermite Model

The extension of the MultiResolution Hermite Model is expressed by the following formula:

$$H(a, m, \delta, \sigma, p, q, x) = p\{1 + a[(x - m + \delta)^2 - 1]\}e^{-\frac{(x-m)^2}{2\sigma^2}} + q \quad (3.7)$$

The following ranges are used as parameters:

$a = [0 \ 0.8]$ is a parameter that assesses the central deformation of the Gaussian
 $m = [7 \ 50]$ is the mean of the Gaussian and allows to shift the function along the x-axis
 $\delta = [0.1 \ 2]$ is a parameter that assesses the asymmetry of the model
 $\sigma = [1 \ 23]$ is the standard deviation of the Gaussian
 $p = [-150 \ 0]$ is a multiplier parameter that allows us to change the sign of the model
 $q = [0 \ 255]$ is a parameter that allows to shift the function along y-axis
 x is a vector of the same length of the cross-section of the vessel
 $e^{-\frac{(x-m)^2}{2\sigma^2}}$ is the Gaussian function.

Once the model has been built it is necessary to compare it with an intensity profile that has been taken manually.

This has been done using a matlab command “*improfile*” that returns pixel-value cross-sections along line segments, selects equally spaced points along the path we specify and then uses interpolation to find the intensity value for each point. Once we have the intensity profile, we need optimization criteria that give us the best value for each parameters. Therefore we used the PSO (*Particle Swarm Optimization*) algorithm to find the best parameters. An explanation of the algorithm follows.

3.4 Introduction to PSO

3.4.1 Particle Swarm Optimization, an optimization technique

Particle Swarm Optimization (PSO) is an evolutionary computation technique developed by Kennedy and Eberhart in 1995.

The particle swarm concept originated as a simulation of a simplified social system.

The original intent was to graphically simulate the graceful but unpredictable choreography of a bird flock. Initial simulations were modified to incorporate nearest-neighbour velocity matching, eliminate ancillary variables, and incorporate multidimensional search and acceleration by distance.

At some point in the evolution of the algorithm it was realized that the conceptual model was, in fact, an optimizer [15].

The term PSO refers then to a relatively new family of algorithms that may be used to find optimal or near to optimal solutions to numerical and qualitative problems.

Particle Swarm Optimization has roots in two main component methodologies. Perhaps more obvious are its ties to artificial life (A-life) in general, and to bird flocking,

fish schooling, and swarming theory in particular [16].

A number of scientists have created computer simulations of various interpretations of the movement of organisms in a bird flock or fish school.

Reynolds, Heppner and Grenander presented simulations of bird flocking. Both of these scientists had the insight that local processes, such as those modeled by cellular automata, might underlie the unpredictable group dynamics of bird social behaviour.

Both the models relied heavily on manipulation of inter-individual distances; that is, the synchrony of flocking behaviour was thought to be a function of bird's efforts to maintain an optimum distance between themselves and their neighbours.

One reason to develop the simulation was to model human social behaviour, which is of course not identical to fish schooling or bird flocking. One important difference is its abstractness.

Birds and fishes adjust their physical movement to avoid predators, seek food and mates, optimize environmental parameters such as temperature. Humans adjust not only physical movement but cognitive or experiential variables as well.

PSO is similar to a genetic algorithm (GA) in that the system is initialized with a population of random solutions ; it is unlike a GA, however, in that each potential solutions, called *particles*, are then “flown ”through the problem space [15].

It requires only primitive mathematical operators, and is computationally inexpensive in terms of both memory requirements and speed [16].

3.4.2 PSO algorithm

The PSO algorithm is population-based: a set of potential solutions evolves to approach a convenient solution (or set of solutions) for a problem.

Being an optimization method, the aim is finding the global optimum of a real-valued function (*fitness function*) defined in a given space (*search space*).

In the PSO algorithm each individual is called a “particle ”. The search is performed by a population of particles i , each having a location vector $p = (\vec{p}_{i1}, \dots, \vec{p}_{ij}, \dots, \vec{p}_{id})$, and representing a potential solution in a d -dimensional search space.

Each particle i also keeps track of its velocity vector, \vec{v}_i , which determines the direction and how far a particle will move in the next iteration.

The fitness of a particle is determined by an evaluation function $F(\vec{p}_i)$. Particles move through the search space in discrete time steps.

In order to provide a balance between local (with a higher tendency to converge to solution in close vicinity) and global search (looking for overall good solutions) an inertia

weight is used [16].

PSO algorithm:

1. Initialize a population of particles with positions and random velocities on d dimension:

$$p = (\vec{p}_0, \dots, \vec{p}_i, \dots, \vec{p}_N) \text{ and } v = (\vec{v}_0, \dots, \vec{v}_i, \dots, \vec{v}_N)$$

2. Evaluate the desired optimization fitness function for each particle in d variables:

$$F(p) = (F(\vec{p}_0), \dots, F(\vec{p}_i), \dots, F(\vec{p}_N))$$

3. Compare particle's fitness evaluation $pbest$. If current value is better than $pbest$, then set $pbest$ value equal to the current value, and the $pbest$ location equal to the current location in d -dimensional space

4. Compare fitness evaluation with the population's overall previous best. If current value is better than $gbest$, then reset $gbest$ to the current particle's array index and value

5. Modify the particle's velocities based on the previous $pbest$ and $globalbest$ positions:

$$\vec{v}_i^{new} = \omega \vec{v}_i + \varphi_1 (\vec{p}_i^{best} - \vec{p}_i) + \varphi_2 (\vec{p}_g^{best} - \vec{p}_i) \text{ for } 1 \leq i \leq N$$

with::

$$\varphi_1 = c_1 rand()$$

$$\varphi_2 = c_2 rand()$$

$c_1, c_2 =$ constant acceleration factors

$\omega =$ inertia weight

6. Update the particle's locations:

$$\vec{p}_i^{new} = \vec{p}_i + \vec{v}_i^{new} \text{ for } 1 \leq i \leq N$$

7. Terminate if some conditions are met (e.g. a sufficiently good fitness or a maximum number of iterations)

8. Otherwise: Go to step 2.

The acceleration constants c_1 and c_2 represent the weighting of the stochastic acceleration terms that pull each particle toward $pbest$ and $gbest$ positions. Thus an adjustment of these constants changes the amount of "tension" in the system. Low values allow particles to roam far from target regions before being tugged back, while high values result in abrupt movement toward, or past, target regions.

Almost all the applications used to set the accelerations constant c_1 and c_2 equal to 2.0, but we have reduced this value to 1.3.

Furthermore a suitable selection of the inertia weight provides a balance between global and local exploration and exploitation, and results in fewer iterations on average to find a sufficiently optimal solution [15], we have set this value to 0.4.

Particle Swarm Optimization is an extremely simple algorithm which seems to be effective for optimizing a wide range of functions.

It is highly dependent on stochastic processes, like evolutionary programming. The adjustment toward $pbest$ and $gbest$ by the particle swarm optimizer is conceptually similar to the *crossover* operation utilized by genetic algorithms. It uses the concept of *fitness*, as all evolutionary computation paradigms do [15].

Unique to the concept of particle swarm optimization is flying potential solutions through hyperspace accelerating toward "better" solutions.

Generally speaking, Particle Swarm Optimization, like the other evolutionary computation algorithms, can be applied to solve most optimization problems and problems that can be converted into optimization problems.

Among the applications areas, the most potential are: system design, multi-objective, optimization, classification, pattern recognition, biological system modeling, scheduling, signal processing, games, robotic applications, decision making, simulation and identification [15].

3.5 Conclusion

In the present chapter, I showed how I built the Extended MultiResolution Hermit Model: I added three parameters to the previous formula (Wang's Hermite Model) and, for each parameter, I explained its specific function inside the model. The added parameters allow to fit an asymmetric intensity profile; aspect that we are really interested in. In the second part of the chapter, I introduced the algorithm that we used to find the best value for each parameter, the PSO (Particle Swarm Optimization). The PSO was used in previous studies by School of Computing's researchers, and it was proved to be very potent and capable to solve problems with 30 dimensions; for these reasons, we didn't search for other optimization criteria, and we just used the PSO.

Chapter 4

Implementation of Extended MHM and PSO

4.1 Introduction

In the following chapter, I will explain how I extract the intensity retinal vessel profile from an image, how I adapt the parameters of the Extended Hermite Model and how I modify the general PSO code for adapting it to my study.

4.2 Extraction of intensity retinal vessel profile

The image that contains the vessels under analysis is uploaded using the matlab command *imread*; it reads the colour image and returns an array containing the image data. If the file contains a true colour image the returned array is an M-by-N-by-3 array. Comparing the different channels of the original RGB image, it was observed that the blood vessels appeared in highest contrast against the background in the green channel, see Figure 4.2.

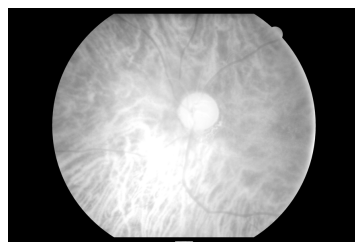
Once the image has been read, using two black points, I identified the vessel sections we are interested in measuring.

By doing this, we are guaranteed to use the same section every time. Thus we obtain 60 sections: 31 arteries' section and 29 veins' section.

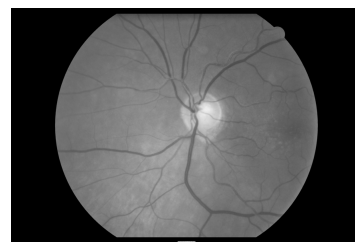
At this point we use the matlab command *improfile* for each section. It consists in selecting a cross section (tracing a line) with the mouse (see Figure 4.3), clicking in two different points, and *improfile* computes the intensity values along the line selected, see Figure 4.3, 4.4. Precisely *improfile* selects equally spaced points along



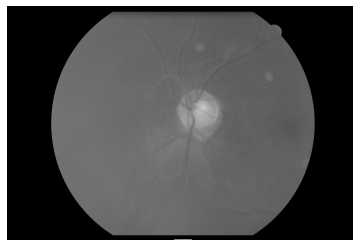
Figure 4.1: RGB image



(a) Red channel



(b) Green channel



(c) Blue channel

Figure 4.2: Extraction of the color channels from an RGB image

the path specified and then uses interpolation to find the intensity value for each point, returning the interpolated data values in a N-by-1 vector.

Final step is to display a plot of the computed intensity values; to be more precise, it creates a two-dimensional plot of intensity values versus the distance along the line segment.

Some typical intensity profiles are shown in Figure 4.4 :

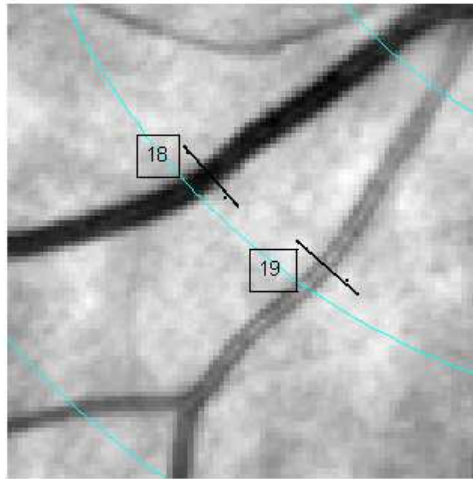
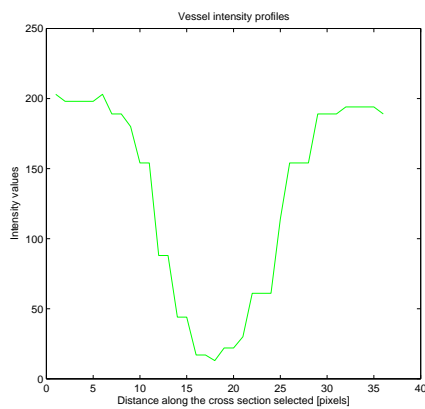
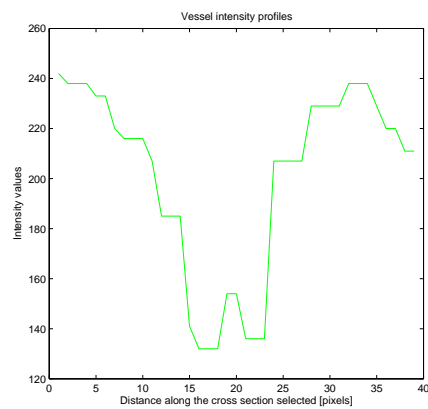


Figure 4.3: Cross section selected with improfile of vessels number 18,19 vein and artery respectively



(a) Manual intensity profile of section 18,
a vein



(b) Manual intensity profile of section 19,
an artery

Figure 4.4: Examples of manual intensity profiles returned using improfile

4.3 Adaption of Extended MHM parameters

The trend of the Extended MultiResolution Hermite Model is controlled by 6 parameters, as explained in the previous chapter.

It is really important to find the best range of values for each parameter, because they control the shape of the profile. We proceed tentatively, fixing 4 parameters and changing the other 2, looking at the respective plot we can see how the parameters control the different shapes of the fitness.

The length of the cross section has been fixed at 70 pixels, due to the fact that the vessel being analyzed are not larger than 50 pixels.

4.3.1 Variation of the parameter a and σ

We change the value of $a = 0, 0.4, 0.6, 0.8$ and $\sigma = 1, 2, 3, 4$; and the values of the other parameters are fixed: $m = 35$, $\delta = 0.01$, $p = -70$ and $q = 255$.

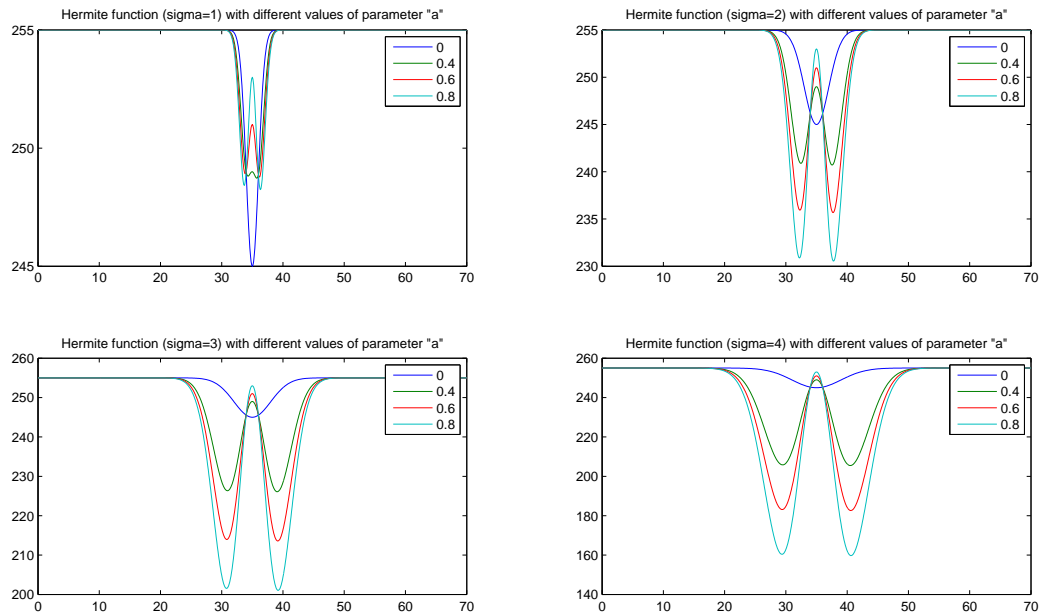


Figure 4.5: Variation of the parameter of a and σ with m , δ , p and q fixed respectively at 35, 0.01, -70 and 255

Looking Figure 4.5 we observe that the curve is centered in m (35^{th} pixel) and it is essentially symmetric due to δ close to 0.

With $a = 0$ the curve is a simple Gaussian whereas the two peaks are further apart

when a increases. Moreover increasing the variance, σ , of the Gaussian the "spread" of the distribution increases around the mean.

4.3.2 Variation of the parameter δ and σ

We change the value of $\delta = 0.01, 0.5, 1, 2$ and $\sigma = 1, 2, 3, 4$; and the values of the other parameter are fixed: $a = 0.6, m = 35, p = -70$ and $q = 255$.

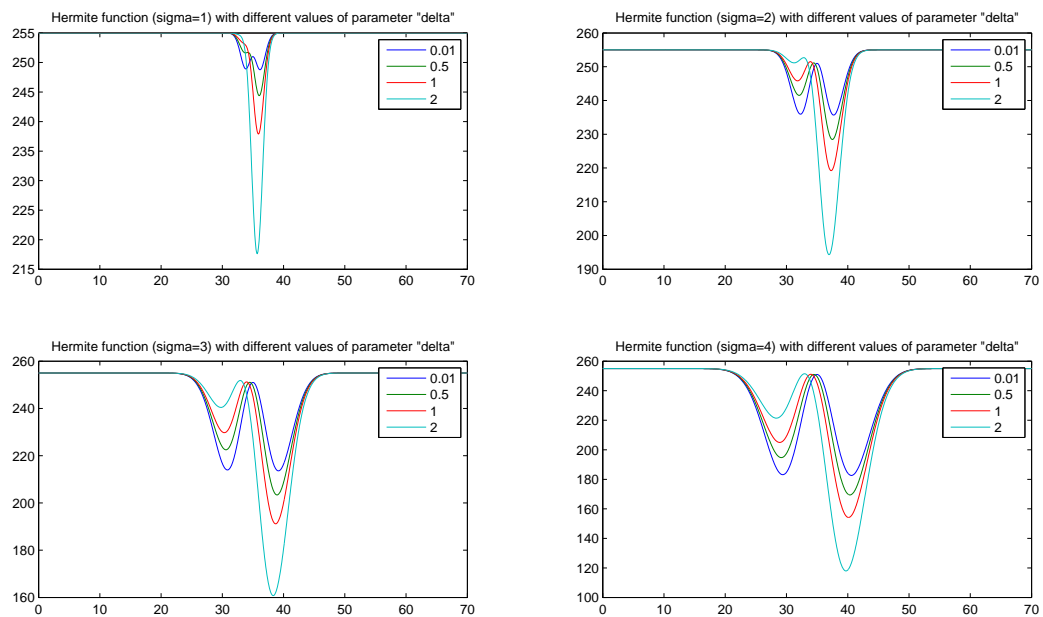


Figure 4.6: Variation of the parameter of δ and σ with a, m, p and q fixed respectively at 0.6, 35, -70 and 255

Looking Figure 4.6 we see that the curve has a deep deformation in the centre, due to $a = 0.6$.

Changing the value of δ we can observe how rapidly the symmetry of the curve changes.

4.3.3 Variation of the parameter m and σ

We change the value of $m = 7, 35, 50, 70$ and $\sigma = 1, 2, 3, 4$; and the values of the other parameter are fixed: $a = 0.4, \delta = 0.01, p = -70$ and $q = 255$.

Looking Figure 4.7 we observe that increasing m from 7 to 70 the curve shifts along the x-axis.

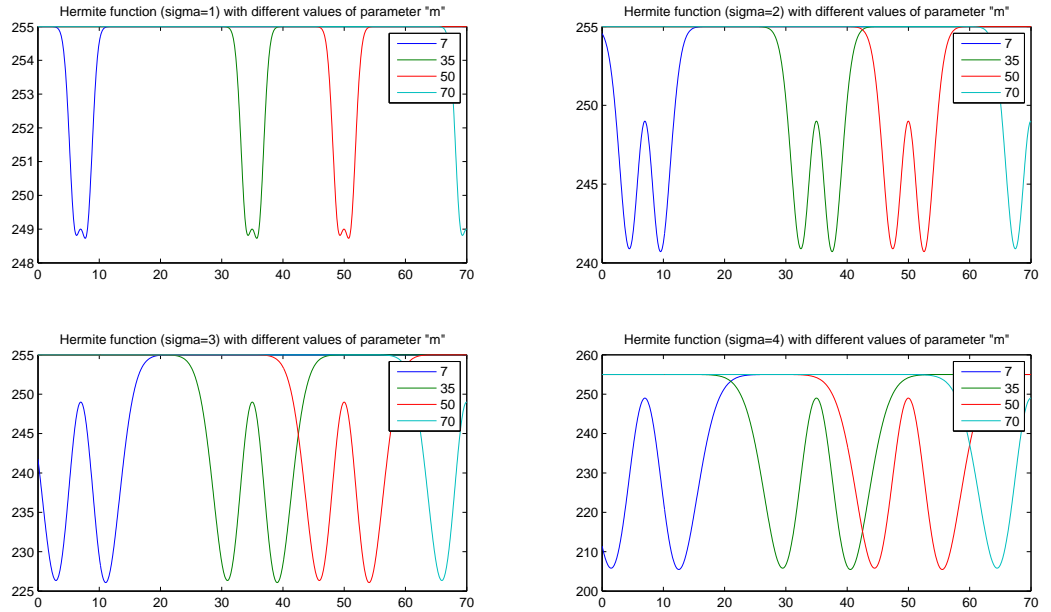


Figure 4.7: Variation of the parameter of m and σ with a , δ , p and q fixed respectively at 0.4, 0.01, -70 and 255

4.3.4 Variation of the parameter a and δ

We change the value of $a = 0, 0.4, 0.6, 0.8$ and $\delta = 0.01, 0.5, 1, 2$; and the values of the other parameter are fixed: $m = 35$, $\sigma = 2$, $p = -70$ and $q = 255$, Figure 4.8.

4.3.5 Variation of the parameter p and σ

We change the value of $p = -150, -100, -10, 1$ and $\sigma = 1, 2, 3, 4$; and the values of the other parameter are fixed: $m = 35$, $\delta = 0.01$, $a = 0.4$ and $q = 255$, Figure 4.9.

Thus we obtain the perfect range of values for acquiring all the possible intensity vessel profiles.

Once we have fixed the best ranges of parameters that allow to cover a broad spectrum of intensity retinal vessel profiles, we have to use an optimizer algorithm that returns the best value of each parameter.

PSO algorithm has been chosen.

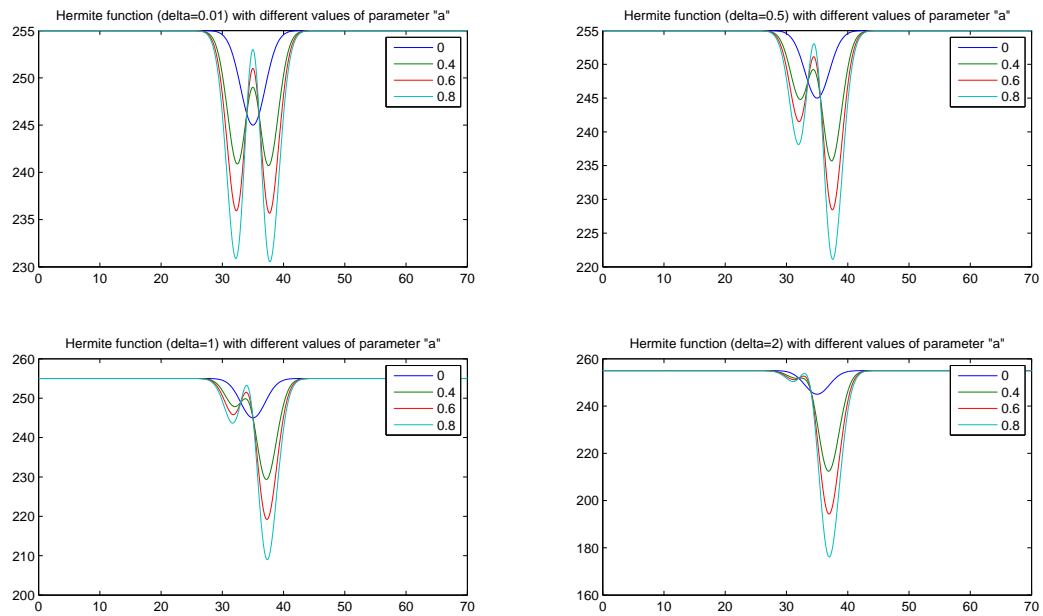


Figure 4.8: Variation of the parameter of a and δ with m , σ , p and q fixed respectively at 35, 2, -70 and 255

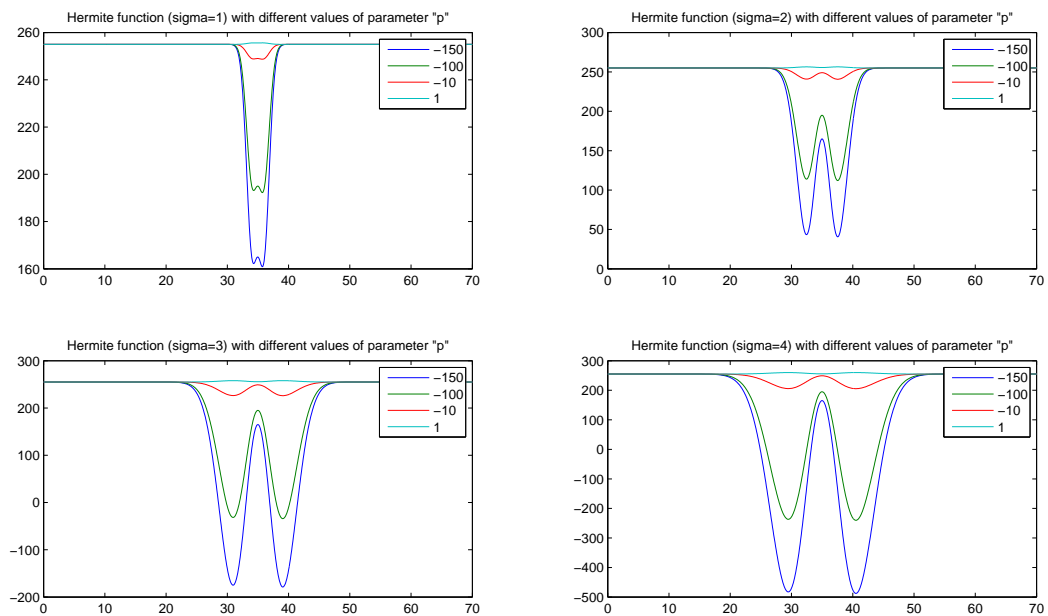


Figure 4.9: Variation of the parameter of p and σ with m , δ , a and q fixed respectively at 35, 0.01, 0.4 and 255

4.4 Adapting PSO (*Particle Swarm Optimization*) algorithm

As discussed in previous chapter, the PSO is an optimization method which needs an error function to optimize.

4.4.1 Error function: Weight function

The error function is calculated by:

$$\overrightarrow{Weight}_i = \frac{1}{N} \sum_{i=1}^H ((improfile - extHermiteFunction_i))^2$$

where:

N = length of the cross section selected using improfile

H = Number of Particles.

Basically it is the difference between the intensity vessel profile achieved with *improfile* and the Extended MHM; done pixel by pixel.

Our goal is to find the minimum value that this function can obtain.

This value will correspond to the fitness function which better represents the characteristics of the intensity profile, taken using improfile.

4.4.2 Initialization of the population of particles

Our 6 parameters are collected in one particle; thus a “*particle*” is defined by:

$$\text{particle}(i) = (a_i, m_i, \delta_i, \sigma_i, p_i, q_i)$$

with i = number of particles

Originally the first particle was built with a mean of every parameter and the other ones were obtained adding gaussian noise (with $SD=1/6(\text{range parameters})$) at the first one, issuing this formula:

$$\text{Particle}(1,:) = [(\min(A)+\max(A))/2, (\min(M)+\max(M))/2, (\min(\text{delta})+\max(\text{delta}))/2, \dots \\ (\min(\text{sigma})+\max(\text{sigma}))/2, (\min(P)+\max(P))/2, (\min(Q)+\max(Q))/2]$$

```
for i = 1 : NUM_Particles
```

```
    Particle(i,:) = Particle(1,)+ rangeVECT/6*randn(1,6)
```

```
end
```


with `rangeVECT` = vector containing the ranges of every parameters.

With this initialization we came upon problems of stability and robustness, due to the possibility of missing best parameters combination.

Thus we decided to initialize the population of particles without any gaussian component.

Consequentially the formula for the particles becomes:

```
Particle(:,1) = linspace(min(A), max(A), NUM_Particles)
Particle(:,2) = linspace(min(M), max(M), NUM_Particles)
Particle(:,3) = linspace(min(DELTA), max(DELTA), NUM_Particles)
Particle(:,4) = linspace(min(SIGMA), max(SIGMA), NUM_Particles)
Particle(:,5) = linspace(min(P), max(P), NUM_Particles)
Particle(:,6) = linspace(min(Q), max(Q), NUM_Particles)
```

Therefore we obtain an uniform distribution of every parameter in the population of particles; it allows to keep the best combination of parameters since the first iteration.

This new formula provides to make the model more robust and consistent.

More accurately the steps of PSO algorithm become:

1. Initialize a population of particles with positions and velocities on 6 dimension:

$$\text{particle} = (\overrightarrow{\text{particle}}_1, \dots, \overrightarrow{\text{particle}}_i, \dots, \overrightarrow{\text{particle}}_N)$$
 and

$$\text{Velocity} = (\overrightarrow{\text{velocity}}_1, \dots, \overrightarrow{\text{velocity}}_i, \overrightarrow{\text{velocity}}_N)$$
 with N = number of particles (100)
2. Evaluate the desired optimization fitness function for each particle in 6 variables:

$$\text{extMHM} = (\text{extMHM}(\overrightarrow{\text{particle}}_1, \dots, \text{extMHM}(\overrightarrow{\text{particle}}_i, \dots, \text{extMHM}(\overrightarrow{\text{particle}}_N)$$
 more accurately

$$\text{extMHM} = (\text{extMHM}(a_1, m_1, \delta_1, \sigma_1, p_1, q_1), \dots, \text{extMHM}(a_N, m_N, \delta_N, \sigma_N, p_N, q_N))$$
3. Compare particles' fitness evaluation *Weight*. If current value is better than *BestLocalWeight* (error function) then set *BestLocalWeight* to the current value, and the *BestLocalParticle* location equal to the current location in 6-dimensional space

```
for i = 1:NUM_PARTICLES
    if (Weight(i) < BestLocalWeights(i))
        BestLocalWeights(i) = Weight(i);
```

```

        BestLocalParticle(i,:) = particle(i,:);
    end
end

```

4. Compare fitness evaluation with the population's overall best. If the current value (*BestCurrentWeight*) is better than *BestGlobalWeight*, then reset *BestGlobalWeight* to the current particle's array index and value

```

for i = 1:NUM_PARTICLES

    [BestCurrentWeight,bestIndex] = populationBest(Weight);
    if (BestCurrentWeight < BestGlobalWeight)
        BestGlobalWeight = BestCurrentWeight;
        BestGlobalParticle(1,:) = BestLocalParticle(bestIndex,:);
    end
end

```

5. Modify the particle's velocity based on the previous *BestLocalParticle* and *BestGlobalParticle* positions:

$$\vec{Velocity}_i^{\text{new}} = \text{inertia}\vec{Velocity}_i + c_1\text{rand}()\overrightarrow{(\text{BestLocalParticle}_i - \text{particle}_i)} + c_2\text{rand}()\overrightarrow{(\text{BestGlobalParticle}_i - \text{particle}_i)}$$

with $1 \leq i \leq \text{NUMPARTICLES}$

In matlab code, this has been implemented with the following code:

```

for i = 1:NUM_PARTICLES
    C1 = 1.3*rand(1,6);
    C2 = 1.3*rand(1,6);

    Velocity(i,:) = inertia*Velocity(i,:)+...
        C1.*(BestLocalParticle(i,:)-particle(i,:))+...
        C2.*(BestGlobalParticle-particle(i,:));
end

```

6. Update the particle's locations:

$$\overrightarrow{particle}_i^{\text{new}} = \overrightarrow{particle}_i + \overrightarrow{Velocity}_i^{\text{new}}$$

with $1 \leq i \leq \text{NUMPARTICLES}$

7. Terminate if some conditions are met (eg. a sufficiently good fitness or a maximum number of iterations)

8. Otherwise: Go to step 2.

Relevant importance in PSO algorithm has been taken by the number of particles and number of iterations; due to the fact that identifying the essential number of iterations allows to decrease the computational time.

4.4.3 Identification of the number of iterations

The sufficient number of iterations can be gathered from an analysis of the error function (*BestGlobalWeight*).

At the first iteration the error function is really important and going on with the iterations we observe that its value decreases due to the fact that the PSO is trying to find the best values for the parameters, moving the fitness of the Extended MHM towards the manual intensity vessel profile.

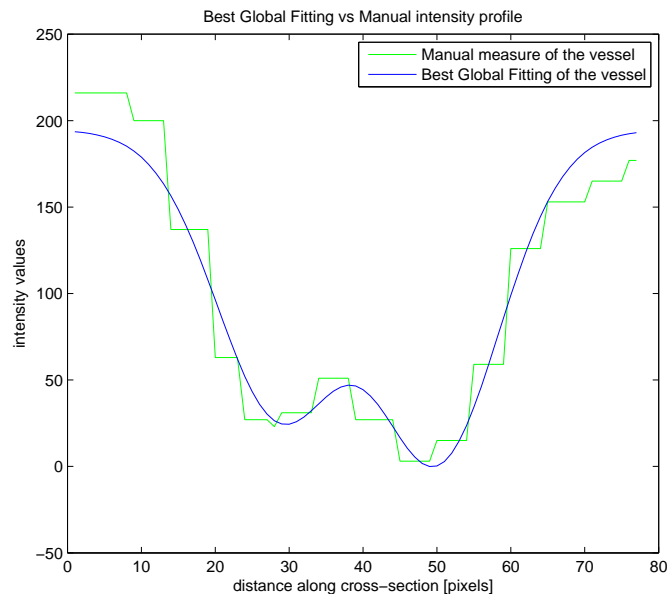


Figure 4.10: Best Fitting obtained running the algorithm for 130 iterations

A characteristical trend of values of the *BestGlobalWeight* follows:

```

Best_GlobalWeight is 11113.1, Entering iteration 1
Best_GlobalWeight is 4879.58, Entering iteration 2
Best_GlobalWeight is 4398.8, Entering iteration 3
Best_GlobalWeight is 2985.46, Entering iteration 4
....
Best_GlobalWeight is 1324.36, Entering iteration 13
Best_GlobalWeight is 1291.42, Entering iteration 14
Best_GlobalWeight is 1270.83, Entering iteration 15
Best_GlobalWeight is 1191.93, Entering iteration 16
.....
Best_GlobalWeight is 312.21, Entering iteration 79
.....
Best_GlobalWeight is 312.21, Entering iteration 130
    
```

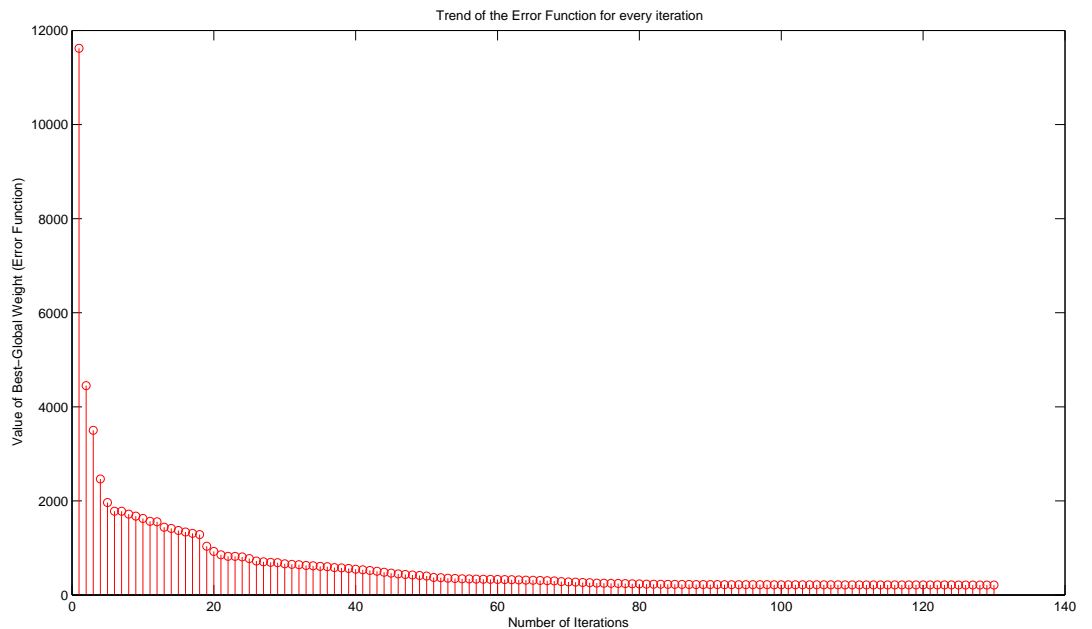


Figure 4.11: Trend of the BestGlobalWeight for 130 iterations

In Figure 4.11 is illustrated the best fitting obtained with 150 iterations; but if we analyze the trend of the *BestGlobalWeight* we derive that the best fitting has been found

at the iteration number 79. Moreover we notice that there's a big gap from the first iteration and the one number 12, afterward the decrease is slower and the convergence at the best fitting is found around 70th iteration.

Investigating most of the sections we decide to fix the number of iterations at 80, in fact further the 80th iteration the *BestGlobalWeight* doesn't decrease more.

Reducing the number of iterations we de-escalate the computational time too. Example: with 130 iterations the algorithm needs around 4 seconds for finding the best fit; instead with 80 iterations it uses 2.5 seconds for find the best fitness.

4.4.4 Values returned by PSO

Finally PSO returns the *bestGlobalParticle* corresponding to the best fitness of the Extended Multiresolution Hermite Model.

Then we can achieve the *bestGlobalFitting* through the Equation:

```
bestGlobalFitting = extMHM(bestGlobalParticle(1, 1), bestGlobalParticle(1, 2), ...
bestGlobalParticle(1, 3), bestGlobalParticle(1, 4), ...
bestGlobalParticle(1, 5), bestGlobalParticle(1, 6), x)
```

Some characteristical fittings follow:

Vessel number 20: Vein with central reflex, Figure 4.12.

Vessel number 3: Vein without central reflex, Figure 4.13.

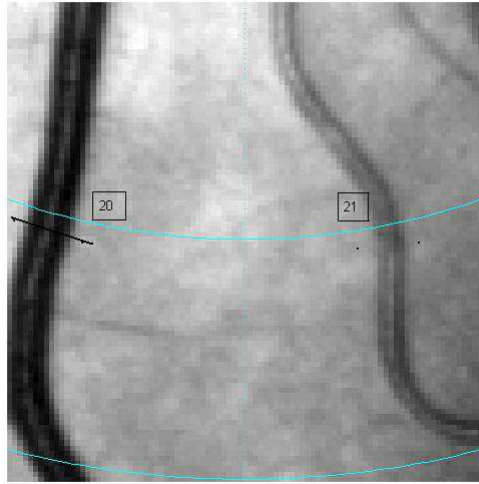
-The typical motion of the particles is like birds flocked around a "roost", a position on the pixel screen that attracted them until they finally landed there [15]. A representation is given in Figure 4.14.

Once that the best fitness has been obtained, the next goal is to find an appropriate algorithm for measure the vessel's width.

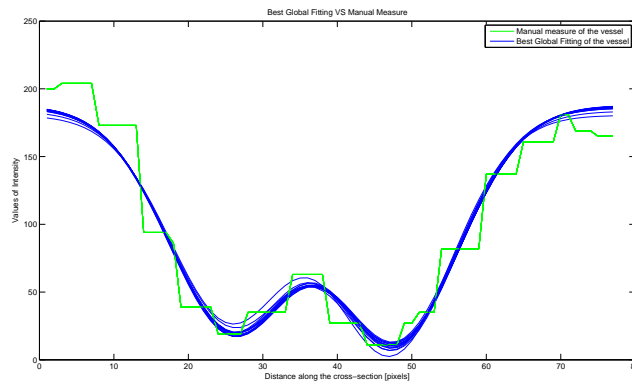
4.5 Conclusions

In the first part of the present chapter, I explained how I extracted the retinal vessel profile from an image, using the matlab commands *imread* and *improfile*. Then, an experimental part in which I studied the best ranges for every parameter, a , m , δ , σ , p and q , in order to cover a broad spectrum of different intensity profiles.

The final part of the chapter regards the adaptation of the general PSO's code to my study: I derived an error function to be minimized, I initialized the population of

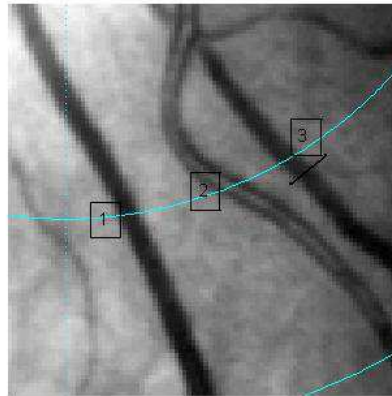


(a) Image containing vessel number 20

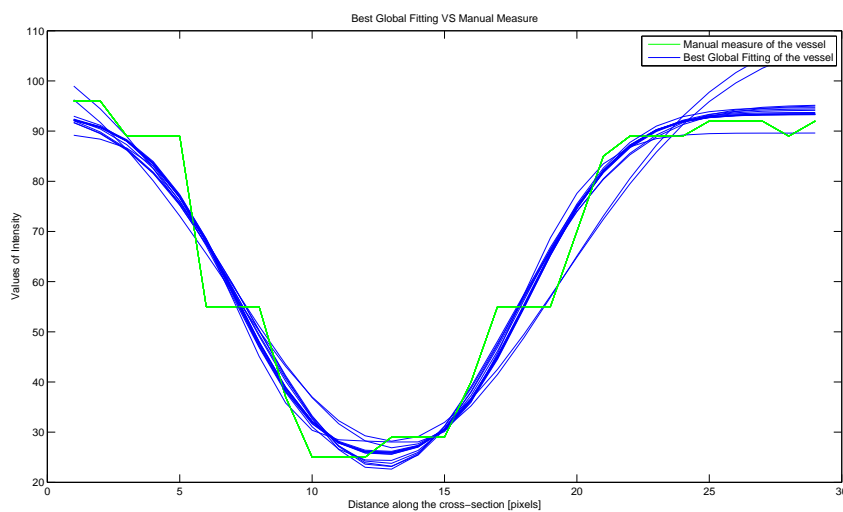


(b) Trend of the BestGlobalfitting for a vein (N.20) with central reflex and its intensity profile taking manually

Figure 4.12: Image containing vein N.20 and the fitness obtained with the model compared with the manual one



(a) Image containing vessel number 3



(b) Trend of the BestGlobalfitting for a vein (N.3) without central reflex and its intensity profile taking manually

Figure 4.13: Image containing vein N.3 and the fitness obtained with the model compared with the manual one

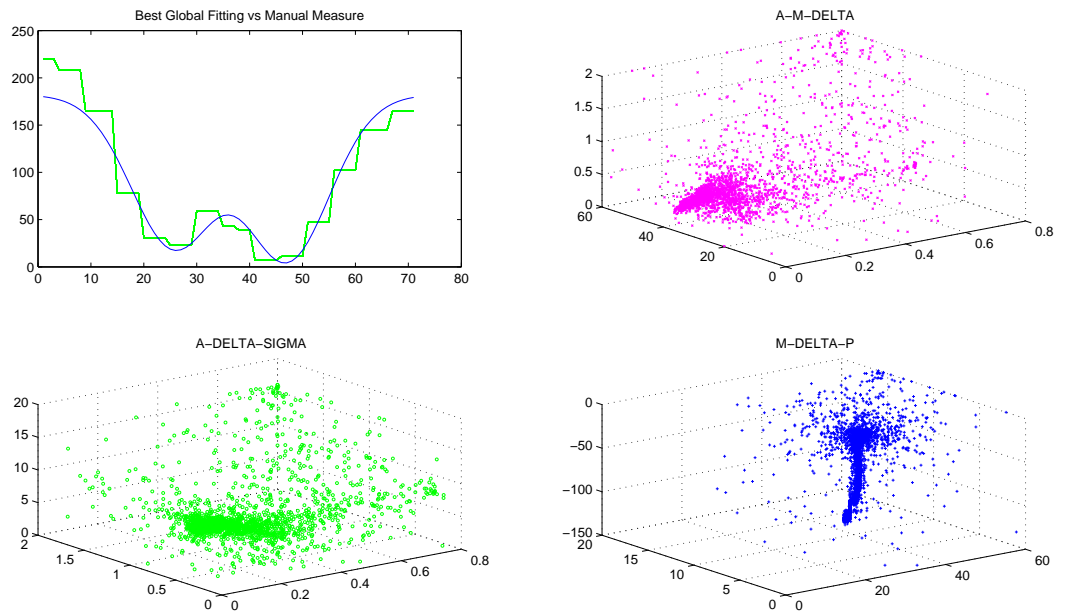


Figure 4.14: Motion of 100 particles during 80 iterations, they start in random position and after some iterations they converge at the best values

particles, removing all the gaussian components and, finally, I identified the minimum number of iterations required to find the best value for each parameter.

Chapter 5

Implementation of algorithms for measuring vessel diameter

5.1 Introduction

In the 2nd chapter I have listed many algorithms that were used for measuring the retinal vessel's width in previous studies. In the present chapter, I will describe the flex technique, shown during the validation phase, to underestimate the width; then I explain how I implement the FWHM, the algorithm that we have finally chosen for this purpose.

5.2 Implementation of flex points technique

Previously Hunter et al. have used a simple line search procedure to locate a zero of the second derivative of their model (2D Gaussian Model), and the offset of this inflection point defines the vessel width [7]; because the inflection point of the curve could correspond to the real borders of the vessel.

In order to find the flex points of our Extended MHM model, we have used the central difference method for finding an approximation of the second derivative of the function (*bestGlobalFitting*), and after that we have searched the reversal of sign inside this function.

We should find two or four changes of sign in the second derivative: two changes of sign mean that the *bestGlobalFitting* is a simple Gaussian, and four changes of sign mean that the *bestGlobalFitting* has a shape different from a simple Gaussian, for instance a Double Gaussian or a Hermite function. We will calculate the vessel width as the

distance along the x-axis between the last flex point and the first one.

These steps are implemented in matlab as it follows:

```

for i = 3:(length(x)-3)
dersec(i)=bestGlobalfitting(1,i+1)-2*bestGlobalfitting(1,i)+bestGlobalfitting(1,i-1)
end

for j = 1:length(dersec)-1
    if (dersec(j)<0 && dersec(j+1)>0) || (dersec(j)>0 && dersec(j+1)<0)
        flesso = [j, dersec(j)];
        flex_point = [flex_point; flesso];
    end
end

ind_flex = [flex_point(1,1);flex_point(length(flex_point),1)];
width_flessi = ind_flex(2)-ind_flex(1);

```

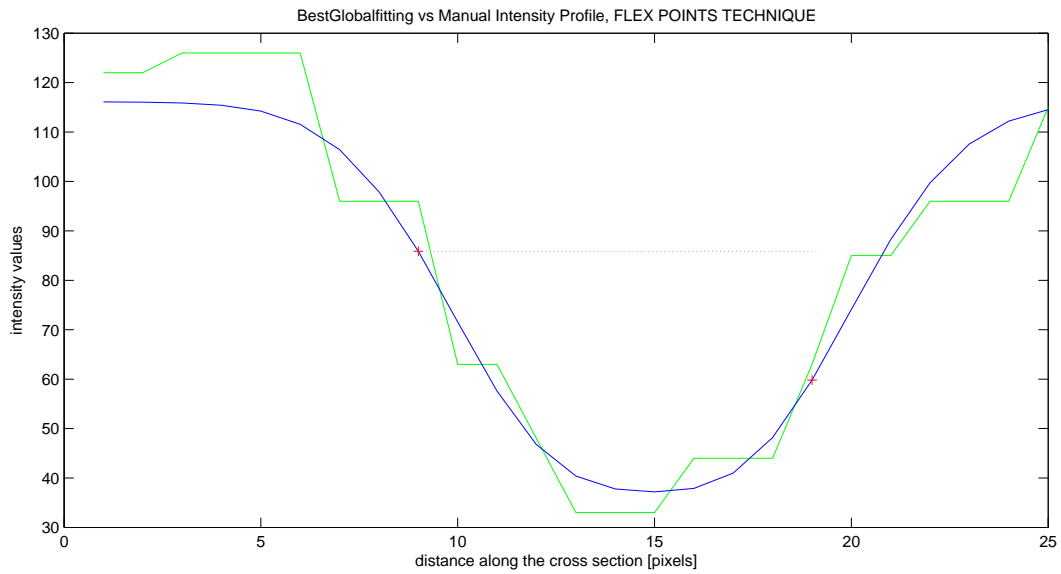
Some graphical representations are shown in Figure 5.1: the first figure (a) shows a simple Gaussian profile and the two red marks indicate the two flex points; the second figure (b) shows a profile characterized by AR, as in the first figure, the marks indicate the two flex points that we use to calculate the vessel width.

5.3 Implementation of FWHM algorithm

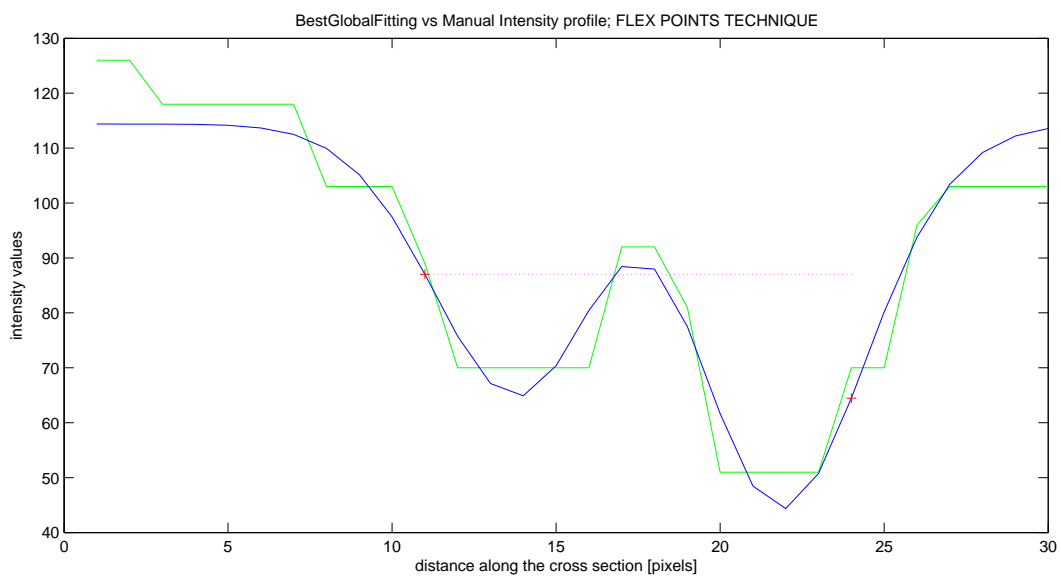
We decided to implement the FWHM (*Full Width Half Maximum*) method due to the better characteristics as compared to other techniques.

This measurement is known from optical theory not to have been largely influenced by aberrations and focus errors, and measured width does not depend on the intensity of the light streak. Another important aspect is that the FWHM algorithm is not sensitive to the presence/absence of an AR or to different background levels; then it can be applied to all the intensity profiles without making distinctions.

As explained in the 2nd chapter, the Full Width Half Maximum (FWHM) is an expression of the extent of a function, given by the difference of the extreme values of



(a) A typical vessel profile without central light, a simple Gaussian profile



(b) A typical vessel profile with central light

Figure 5.1: Two examples of typical vessel profiles, flex points technique is applied to measure the vessel width

the independent variable at which the dependent variable is equal to half of its maximum value.

Then this approach calculates the “half height point ”on the left and right sides of the initial estimated midpoint of the profile. On each side, the minimum and maximum

intensity levels are calculated, and the “half height point ”is located where the profile crosses the mid point in intensity between the minimum and maximum. The FWHM estimate of the profile width is then the distance between these half height points; this approach is also called half-height at full width (HHFW).

Then it is required to find 4 values:

- The MinLeftLevel
- The MinRightLevel
- The MaxLeftLevel
- The MaxRightLevel

5.3.1 Search of maximum values

There could be two different cases:

- begin and the end of the fitness could be the maximum values of the fit
- maximum values could be located along the curve

To cope with this aspect I follow these steps:

1. I initialize the two maximum values (MaxLeftLevel and MaxRightLevel) with the first and last value of the fitness respectively

```
MaxLeftLevel = [1 bestGlobalFitting(1,1)]
MaxRightLevel = [ length(x) bestGloxbalFitting(1,length(x))]
(x = length(bestGlobalFitting))
```

2. It is important to check if these values are really the maximum of the fitness (first case) or not.

Analyzing the MaxLeftLevel: if its following value is smaller than it, it means that the first value of the fitness is really the Left Maximum; instead, if the following value is bigger this means that there a Left Maximum along the curve and it has to be found.

We have to proceed in this way: we run along the curve till we will find a value (*firstmax*) for which its previous value and its following one are both smaller; then *firstmax* will be MaxLeftValue.

```

if bestGlobalFitting(1,i) > bestGlobalFitting(1,1)
    if best_Globalfitting(1,i-1) < best_Globalfitting(1,i) && ...
        best_Globalfitting(1,i+1) < best_Globalfitting(1,i)
        MaxLeftLevel = [i best_Globalfitting(1,i)];
    end
end

```

3. Approximately the same procedure is followed to find the MaxRightLevel: if its previous value is smaller than it, it means that it is actually the Right Maximum; instead, if the previous value is bigger we have to run the curve searching for the maximum value, in the same way of the previous step.

```

if bestGlobalFitting(1,i) > bestGlobalFitting(1,x)
    if bestGlobalFitting(1,i-1) < bestGlobalFitting(1,i) &&...
        bestGlobalFitting(1,i+1) < bestGlobalFitting(1,i)
        MaxRightLevel = [i best_Globalfitting(1,i)];
    end
end

```

5.3.2 Search of minimum values

The fitness of the model could have one minimum value or two minimum values. I initialize the vector of the minimum values with an empty vector, and then I run the curve to find the values that verify the condition of minimum.

The minimum value is found when, looking at one value, its previous value and its following one are bigger than it.

```

for i = 2:(length(x)-1)
    if bestGlobalFitting(1,i-1) > bestGlobalFitting(1,i) &&...
        bestGlobalFitting(1,i+1) > bestGlobalFitting(1,i)
        minimo = [i,bestGlobalFitting(1,i)];
        min_fit = [min_fit;minimo];
    end
end

```

The first value to be found is the MinLeftLevel and the second one is the Min-RightLevel.

If the search of the minimum returns only a single minimum, this means that the intensity vessel profile has not the central reflex, and the profile is a shape similar to a Gaussian. In this case I copy the only value found in the second place of the vector of the minimums, in order to have always two minimums, which can have the same value (when the curve has a single minimum) or two different values (when the curve has two distinct minimums).

Once that *MaxLeftLevel*, *MaxRightLevel*, *MinLeftLevel* and *MinRightLevel* have been found, the half heights (left and right) are found with the following formulas:

$$\text{halfminimumleft} = \frac{\text{maxleft} + \text{minleft}}{2}$$

$$\text{halfminimumright} = \frac{\text{maxright} + \text{minright}}{2}$$

Finally the width of the vessel is calculated like the distance between the two half heights.

We have to find the abscissa value in the fit corresponding to the *half minimum left* and *half minimum right*, we implement the following formula:

$$x = x_1 + \frac{(x_2 - x_1)(y - y_1)}{y_2 - y_1}$$

Then in matlab the two abscissas are given by:

```

for i = 2 : length(x)-1
    if (best_Globalfitting(1,i-1) >= half_minimum_left) &&...
        (best_Globalfitting(1,i) <= half_minimum_left)
        indexA = (i-1) + (half_minimum_left-best_Globalfitting(1,i-1))*...
            ((i-(i-1))/(best_Globalfitting(1,i)-best_Globalfitting(1,i-1)));
    end

    if (best_Globalfitting(1,i-1) <= half_minimum_right) &&...
        (best_Globalfitting(1,i) >= half_minimum_right)
        indexB = (i-1) + (half_minimum_right-best_Globalfitting(1,i-1))*...
            ((i-(i-1))/(best_Globalfitting(1,i)-best_Globalfitting(1,i-1)));
    end
end

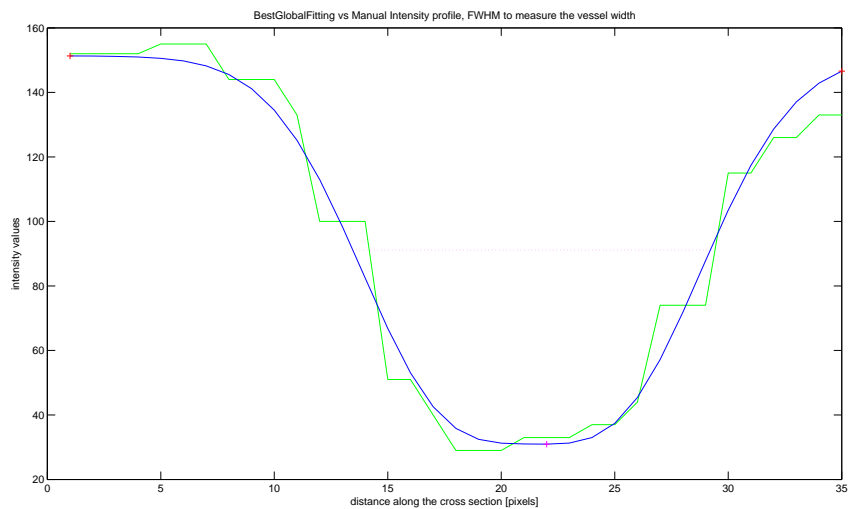
end

```

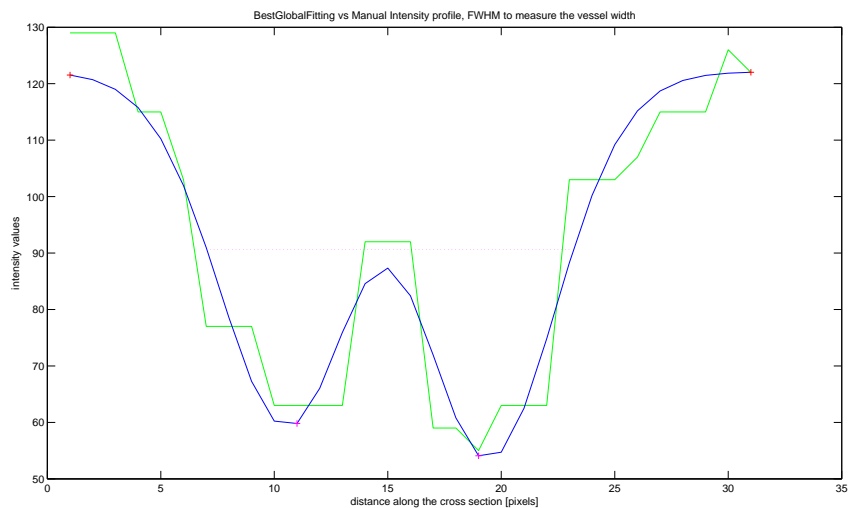
And finally the width of the vessel is calculated with:

$$\text{width} = \text{indexB} - \text{indexA}$$

See Figure 5.2.



(a) A typical vessel profile without central light, a simple Gaussian profile



(b) A typical vessel profile with central light

Figure 5.2: Two examples of typical vessel profiles; in green it is the manual profile, in blue it is the profile obtained with the extended MHM, and in red it is the estimated vessel width

5.4 Conclusions

In the first part of the present chapter I explained how I implemented the flex points technique: I used the central difference to approximate the second derivative of the model's function and finally I searched for the reversal of sign of this approximation function. In the second part I implemented the FWHM algorithm: I searched for the maximum and minimum values of the model's function and then I derived the width at half height between the minimum values and the maximum values.

Chapter 6

Validation of the Extended MHM

6.1 Introduction

In the present chapter I will deal with the validation of the model. In the first part I will show the good properties of the extended MHM. Afterwards I will explain how we tackle the validation: generation of a ground truth and statistical analysis of the data collected. Finally a discussion about the results given by the analysis follows.

6.2 Testing accuracy and robustness of the model

By having the full algorithm we can now find the best fit for each section of the vessels and we would be able to calculate the respective width.

This would be achieved in the following way: we would run the algorithm 15 times in order to observe the stability and accuracy of the algorithm. We decided to run the algorithm for 15 times because it is a good compromise between performance times and statistical significance.

We then load the fundus images containing the vessels, and then select the section which we are interested in measuring the width and let the algorithm run for 15 times. We then compare the 15 results with the same manual intensity profile as used previously.

These are a few examples of the fitness results:

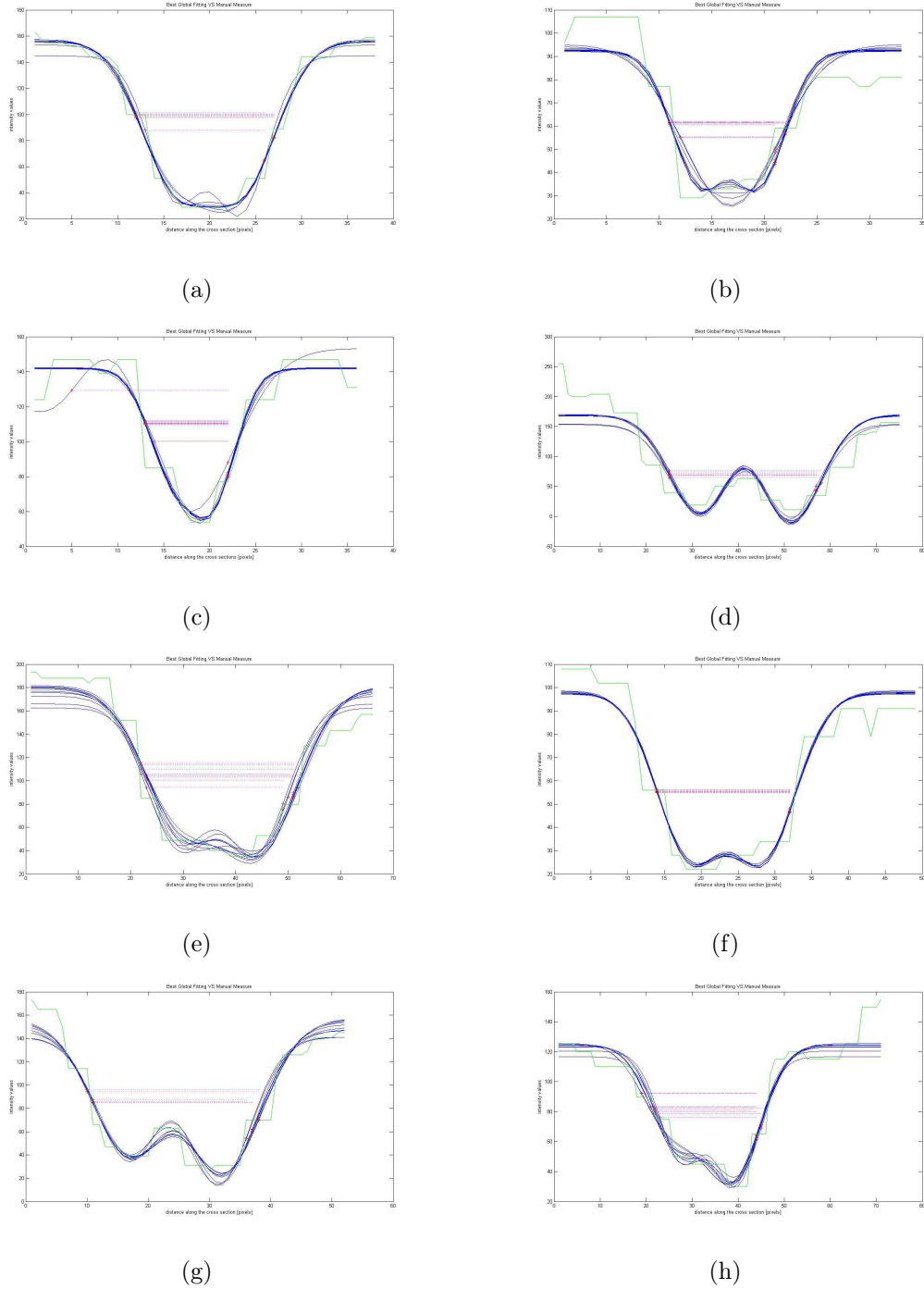


Figure 6.1: Examples of 15 fitnesses regarding the same section, the vessel widths are obtained with flex points method

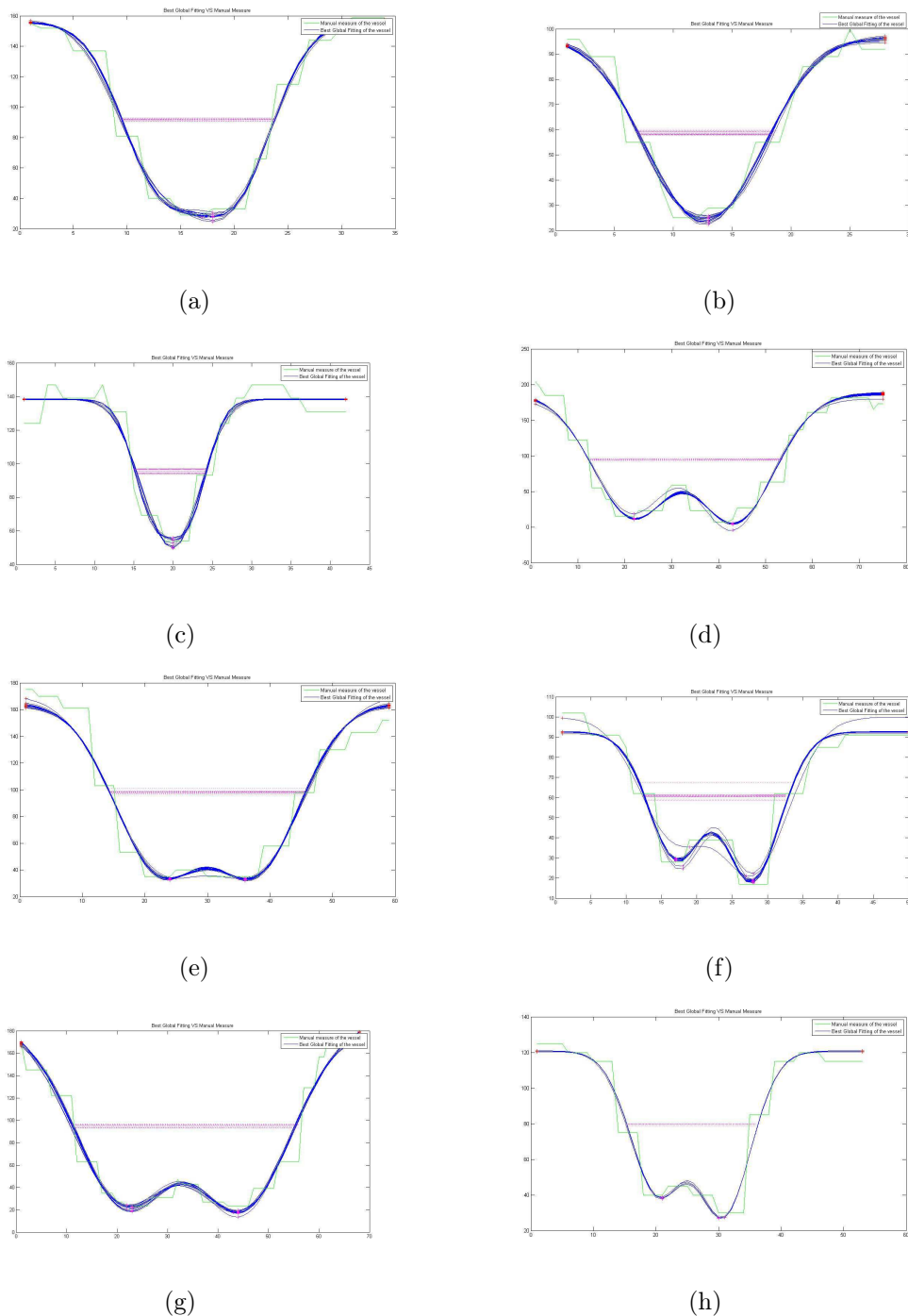


Figure 6.2: Examples of 15 fitnesses regarding the same section, the vessel widths are obtained with FWHM algorithm

Looking at the previous fitnesses, Figure 6.1 and Figure 6.2, it is possible to observe that the extended MHM is accurate and stable enough to be able to follow almost perfectly the manual intensity profile. These positive results are repeatable for 15 times without major alterations.

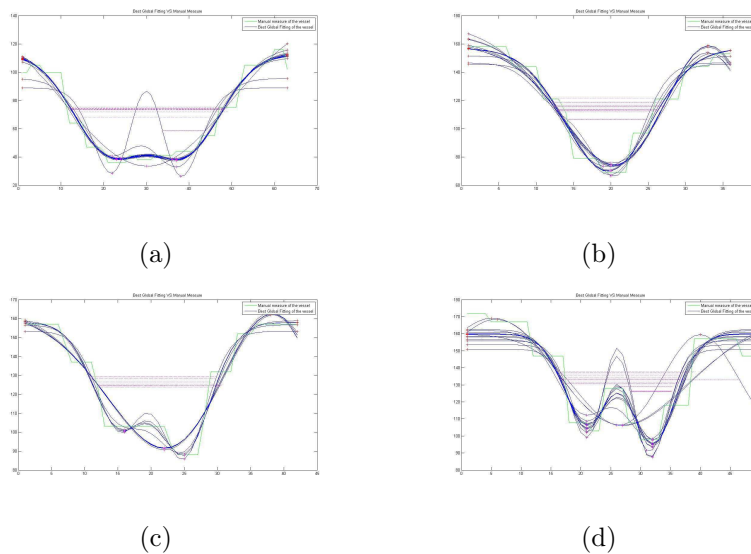


Figure 6.3: Examples of 15 fitnesses regarding the same section, the vessel widths are obtained using FWHM algorithm, the fitness are not good enough

However, it has been observed that fitting sixty vessels, the algorithm showed not follow very well some vessels, in particular in the images that do not present a sensitive intensity difference between the vessel and the background. Such exception happened only a few times, so the algorithm can still be considered stable and accurate.

In Figure 6.3 and Figure 6.4 some wrong fits are pictured ; we observe that running the algorithm for fifteen times for each section it is out of line with the correct fit only a few times. This allows to deem the algorithm efficient and it also triggers future work, in order to make the algorithm applicable to every vessel.

To observe the fifteen widths regarding the same vessel's section we use the matlab's command *hist* and the peak of the histogram acts as the true width of each section. Below are some examples, Figure 6.5:

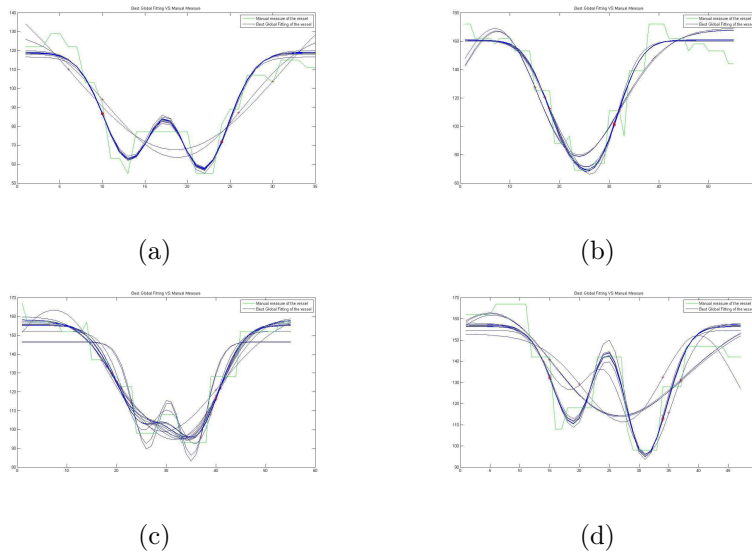


Figure 6.4: Examples of 15 fitnesses regarding the same section, the vessel widths are obtained using flex point technique, the fitness are not good enough

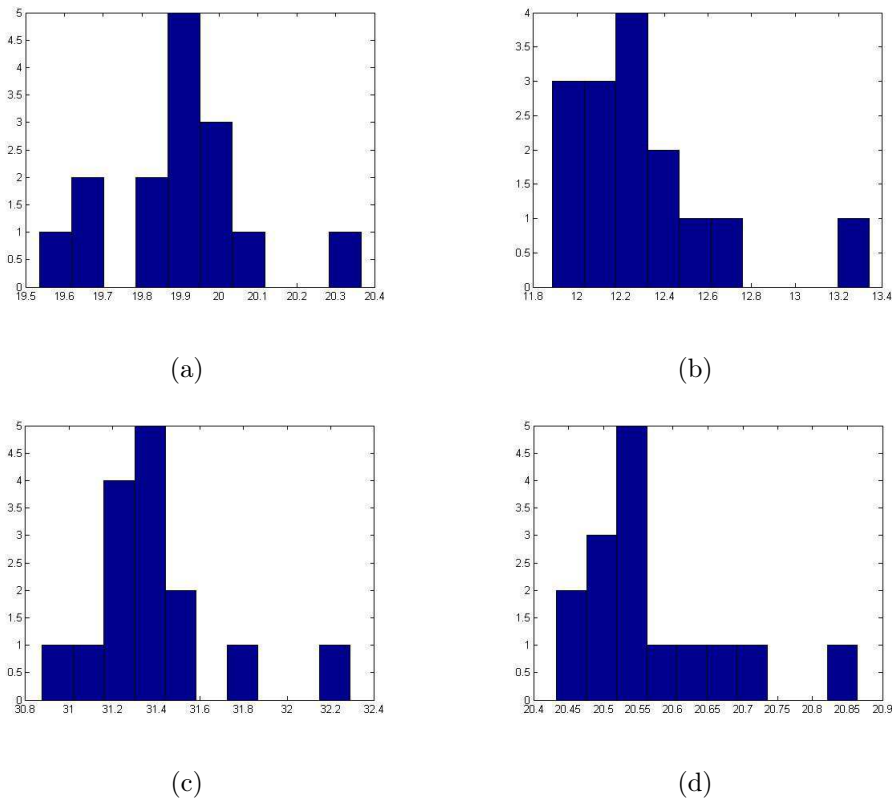


Figure 6.5: Examples of histograms, each one showing 15 widths of the same vessel's section

6.3 Generating the Ground Truth

The first important step needed to validate a model is to generate a ground truth.

In the ground truth are collected all the results obtained through the manual and the automatic measures.

The image set that was provided to us is composed of 60 vessels: 31 arteries and 29 veins. The purpose is to generate a pilot set of manual measurements of vessel widths against which to validate the output of software, finding the width of retinal vessel at well-defined positions.

6.3.1 Observers' measures: protocol used

Three observers participated at this exercise: Paola Nicoli, Emanuele Trucco (professor of Computational Vision at Dundee University) and Peter Wilson (doctor at the Ninewells Hospital of Dundee, registrar in Ophthalmology).

General instructions were given to all three observers: it was requested to repeat the procedure below three times, in three sessions, leaving some time between sessions, so we could have more accurate results.

1. Open Paint (it is found in all versions of Windows and it is easy to use) and load the first image.
2. Identify the cross sections to be taken: each one is indicated clearly by a boxed number and by two black points, on the opposite sides of the vessel and slightly outside the vessel.
3. For each cross-section in the image, position the cursor on the vessel border point closer to the leftmost black point. Record the vessel number and the co-ordinates of this point.
 - The borders of the vessel are not perfectly clear, this is the reason why three different estimates are needed, every observer can follow their own definition of "vessel border ". Every observer has to be consistent with the own definition till the end of the experiment.
4. Move the cursor across the vessel, perpendicularly to its axis, and place the cursor on the opposite vessel border point. Record the co-ordinate of this point.
5. Once processed all the cross-sections in one image, close it and repeat from step 1 for the next image. Do it for all the images.

Then we have three measures for each observer regarding the same cross-section; in total we should have nine observers measurements for each section.

6.3.2 Software's measures: protocol used

The same image set is used: 60 vessels, 31 arteries and 29 veins; the vessels are measured using both the algorithms that we have implemented, flex points and FWHM. Such as the observers' measures, the procedure has to be performed three times, in three different sessions. The protocol used to collect the automatic measurements is done accordingly:

1. Open the software and load the first image.
2. Identify the cross sections to be taken: each one is indicated by two black points.
3. For each cross-section in the image, position the cursor closer to the leftmost black point and click it; then move the cursor across the vessel, perpendicularly to its axis, and place the cursor on the opposite side of the vessel; it does not matter how far it is from the border, and click it the second time. In this way a cross section has been selected, tracing a line to identify it.
4. The software runs for 15 times in the selected cross-section, returning then 15 widths. The widths are depicted in a histogram.
5. The width of the current cross-section is taken like the peak of the histogram.
6. Once processed all the cross-sections in one image, repeat from step 1 for the next image. Do it for all images.

Then we have six automatic measures for each section: three are obtained with the flex points technique and three using the FWHM algorithm.

The ground truth is finally composed of 15 columns of 60 measures each: 3 columns concerning Paola's measures, 3 columns concerning Manuel's measures, 3 columns regarding Peter's measures, 3 columns regarding the measures obtained with the software using the flex points and the last 3 columns regarding the measures given using the FWHM method.

A statistical analysis is required because we want to know if the program measurements are significantly different from those taken by the observer. Our hope is that program measurements are very similar to those taken manually.

6.4 Statistical analysis

In clinical measurement comparison of a new measurement technique with an established one is often needed to see whether they match sufficiently for the new one to replace the old one. Usually a new method has to be evaluated by comparison with an established technique rather than with the true quantity. If the new method matches sufficiently well with the old one, the old one may be replaced. The correct statistical approach is not obvious [17].

In most research the data analysis involves three major steps, roughly done in this order:

- Cleaning and organizing the data for analysis.
- Describing the data (Descriptive Statistics).
- Testing hypotheses and models (Inferential Statistics).

Descriptive Statistics are used to describe the basic data features in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

With descriptive statistics we simply describe what the data shows. They are used to present quantitative descriptions in a manageable format; descriptive statistics reduce plenty of data into a simpler summary.

Inferential Statistics investigate questions, model and hypotheses. For instance, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study.

We would like to describe a specific characteristic for each of the variables (measures) in our study: the mean and the standard deviation.

The **mean** is probably the most commonly used method of describing central tendency. To compute the mean we add up all the values and we divide by the number of values:

$$\hat{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

The **standard deviation** is an accurate and detailed estimate of dispersion. It shows how much variation there is from the mean. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviations

indicate that the data is spread out over a large range of values:

$$SD = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x})^2}{N - 1}}$$

The standard deviation allows to estimate a confidence interval.

A *confidence interval* is an interval in which a measurement or trial falls corresponding to a given probability. Usually, the confidence interval of interest is symmetrically placed around the mean.

A 95% confidence interval is given by:

$$mean \pm 1.96SD$$

This means that 95% of the values of the variable fall in this range.

With inferential statistics we are trying to reach conclusions that extend beyond the immediate data alone.

Most of the major statistics come from a general family of statistical models known as *General Linear Model*. This include the t-test, ANalysis Of VAriance (*ANOVA*), ANalysis Of COVariance (*ANCOVA*), regression analysis and many of multivariate methods like factors analysis, multidimensional scaling, cluster analysis and so on.

During our validation phase we applied to our data the following statistics: scatter plot, Bland-Altman plot and finally a paired t-test.

6.4.1 Scatter Plot

A **scatter plot** is a type of mathematical diagram using Cartesian coordinates to display values for two variables for a set of data.

The data is displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis, and the value of the other variable determining the position on the vertical axis.

A scatter plot will illustrate the degree of correlation between two variables; a line of best fit (alternatively called “trendline”) can be drawn in order to study the correlation between the variables.

The relative correlation of one characteristic to another one can be seen both from how closely points cluster the line, and the correlation coefficient. Values near 1 imply very high correlation between the characteristics, meaning that a change in one characteristic will be accompanied by change in the other characteristic. Positive correlation means that as one increases, so does the other, and is shown on the Scatter diagram as a line with positive slope. Negative correlation implies that as one characteristic increases, the other decreases, and a negative slope is seen in the Scatter diagram.

6.4.2 Bland-Altman plot

A simple plot of the results of one method against the others, though without a regression line, is a useful start but usually the data points will be clustered near the line, and it will be difficult to assess differences between methods. A plot of the difference between the methods against their mean may be more informative.

The Bland & Altman plot is a statistical method to compare two measurement techniques.

Bland and Altman advocated the use of a graphical method to plot the difference scores of two measurements against the mean for each subject and argued that if the new method agrees sufficiently well with the old one, the old one may be replaced.

Here the idea of agreement plays a crucial role in method comparison studies. The Bland-Altman plot is also useful to reveal a relationship between the differences and the averages, to look for any systemic bias and to identify possible outliers. The plot of difference against mean also allows us to investigate any possible relationship between the measurement error and the true value. We do not know the true value, and the mean of the two measurements is the best estimate we have. It could be a mistake to plot the difference against either value separately because the difference will be related to each, a well-known statistical artefact [17].

How to construct it

Consider a set of n samples (objects of unknown volume). Both assays (for example different methods of volume measurement) are performed on each sample, resulting in $2n$ data points.

Each of the n samples is then represented on the graph by assigning the mean of the two measurements as the abscissa (X-axis) value, and the difference between the two values as the ordinate (Y-axis) value.

Hence the Cartesian coordinates of a given sample S with values of S_1 and S_2 determined by the two assays are:

$$S(x, y) = \left(\frac{S_1 + S_2}{2}, (S_1 - S_2) \right)$$

The Bland-Altman method calculates the mean difference between two methods of measurements (the “bias”) and 95% limits of agreement as the $1.96SD$ (with SD = standard deviation of the mean difference), then the limits of agreement are:

$$BIAS \pm 1.96SD$$

If the differences within the limits of agreement are not clinically important, the

two methods may be used interchangeably.

6.4.3 Paired t-test

Early in this century it was shown by W.S.Gossett, writing under the name of “Student”, that the mean of a sample from a Normal distribution with unknown variance has a distribution that is similar to, but not quite the same as, a Normal distribution. He called it the *t distribution*, and we still refer to it as Student’s *t* distribution, see Figure 6.6.

As the sample size increases the sampling distribution of the mean becomes closer to Normal distribution. We use the *t* distribution for estimation and hypothesis testing relating to the means of one or two samples.

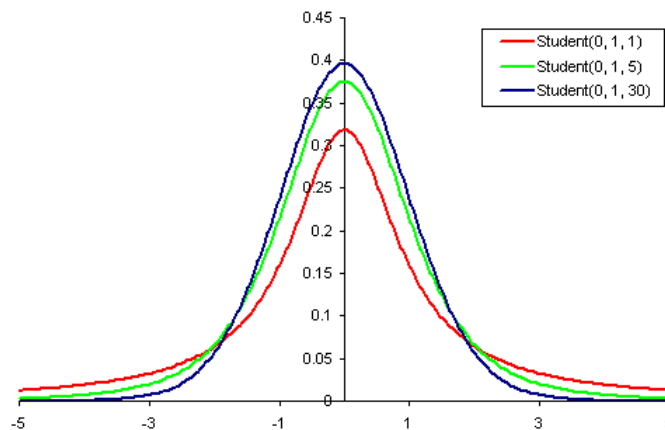


Figure 6.6: Student’s *t* distribution

A *t*-test is any statistical hypothesis test in which the test statistic follows a Student’s distribution if the null hypothesis is true.

The *t*-test assesses whether the means of two groups are statistically different from each other. When we are looking at the differences between scores for two groups, we have to judge the spread of variability of their scores.

Two sample *t*-test for a difference in mean can be either unpaired or paired.

Paired t-test are a form of blocking (it is the arranging of experimental units in groups that are similar to one another), and have greater power than unpaired tests when the paired units are similar.

The **Unpaired t-test** is used when two separate independent and identically distributed samples are obtained, one from each of the two populations being compared.

The t-statistics to test whether the means are different can be calculated as follows. Each of these statistics can be used to carry out either a one-tailed test or a two-tailed test.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{std.error_{(\bar{x}_1 - \bar{x}_2)}}$$

where:

\bar{x}_1 = mean of the first group of measures

\bar{x}_2 = mean of the second group of measures

$std.error_{(\bar{x}_1 - \bar{x}_2)}$ = standard error of the difference between the two means, it is calculated with the following formula:

$$std.error_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{var_1}{N_1} + \frac{var_2}{N_2}}$$

The t-value will be positive if the first mean is larger than the second, and negative if it is smaller.

To test the significance of the hypothesis we need to set a risk level (Alpha level or p-value). In most social research the alpha level is set as 0.05. This probability means that 5 times out of 100 we would find a statistical difference between the means if there was none.

Given the alpha level, the degrees of freedom, and the t-value, we can look the t-value up in a standard table of significance to determine whether the t-value is large enough to be significant. If it is significant we can conclude that the difference between the two means is different.

The *t* distribution has one parameter called *degrees of freedom*. In general the degrees of freedom are calculated as the sample size minus the number of estimated parameters. Hence for a single sample of *n* observations we have *n-1* degrees of freedom.

6.5 Statistical results

The statistical analysis was done by Dr. Weijie Wang of the Institute of Motion Analysis and Research of Ninewells Hospital. All his statistics analysis was carried out using SPSS®(version 17). As the data were normal distributed (skewness value less than twice its standard error) and repeated measures, general linear model, which is an extended version of t-test, was used to test whether there were significant difference from each other. The post-hoc (LCD, The least significant difference) was used to test whether there were differences between subjects or measures, which is equivalent to multiple individual t-test between all pairs of groups. The significant level was set as 0.05.

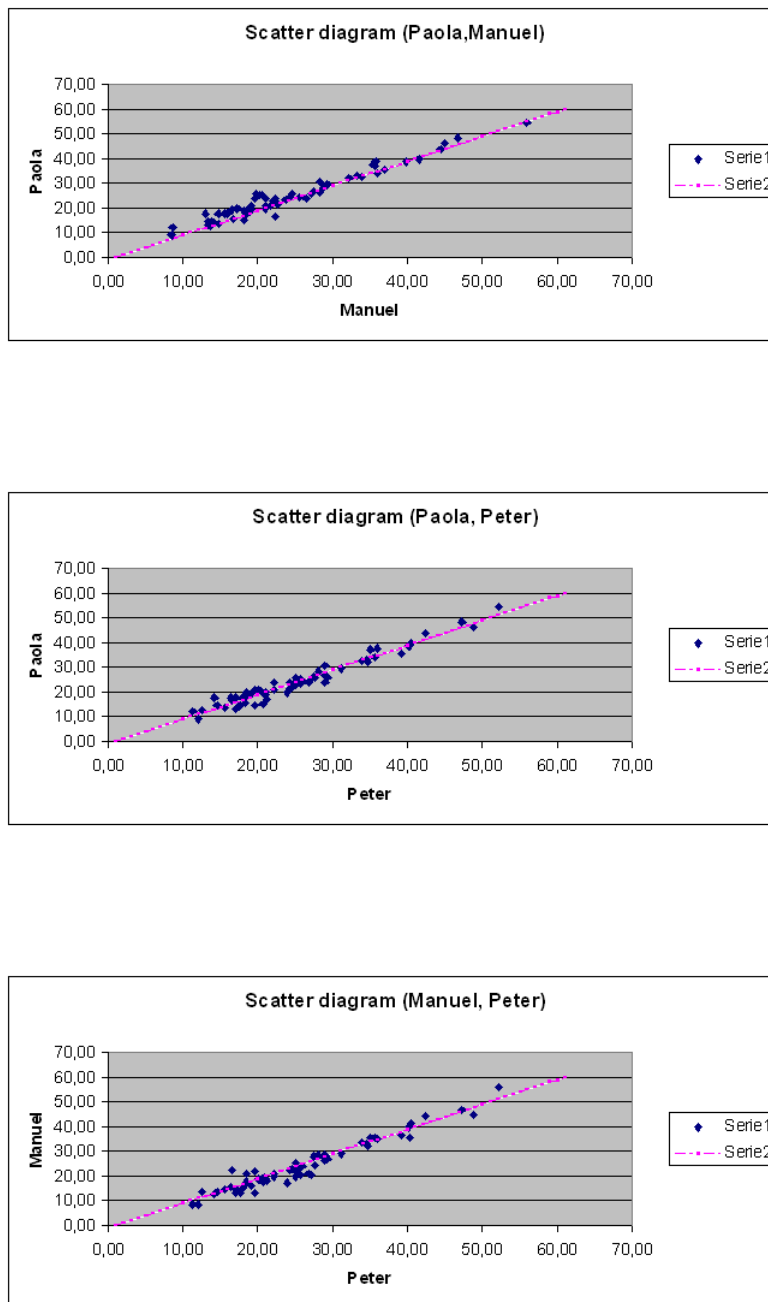


Figure 6.7: Scatter plots, comparison between the three observers, Paola, Manuel and Peter

I also did the statistical analysis using scatter plots, Bland Altman plots and paired t-test. The results match with the ones obtained by Weijie.

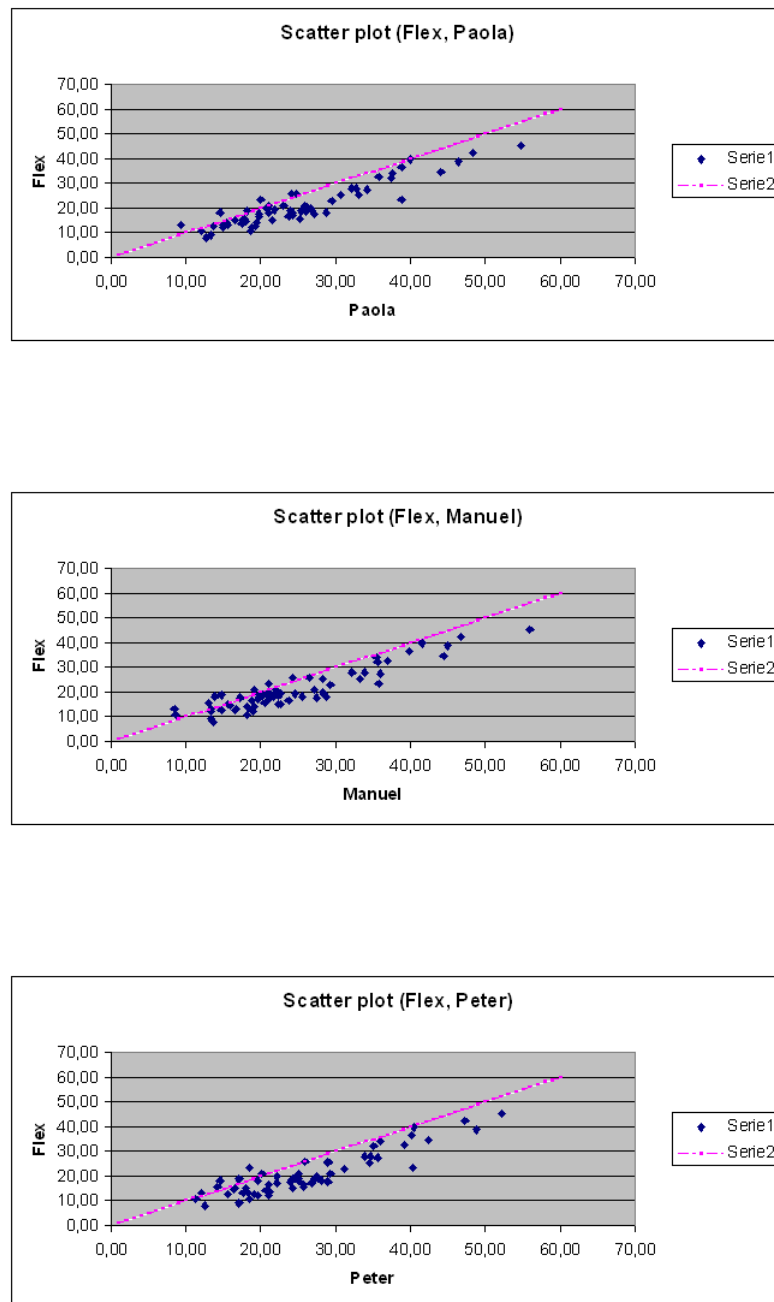


Figure 6.8: Scatter plots, comparison between Flex points technique and the observers

The Scatter plots show that there is no systematic difference between the three observers, Figure 6.7 .

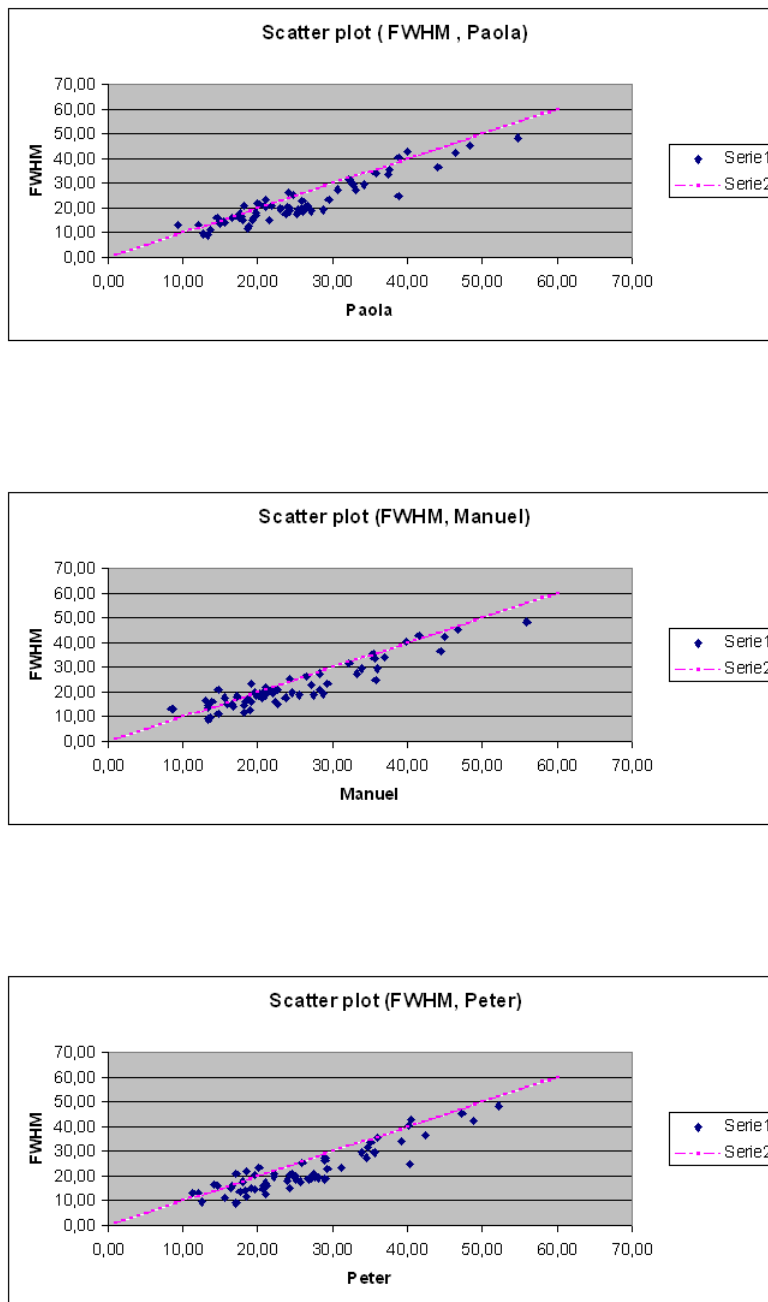


Figure 6.9: Scatter plots, comparison between the FWHM algorithm and the three observers

Looking the plots regarding the flex points technique and the single observers, Figure 6.8, we notice that there is a systemic difference: the flex points technique underestimates

the measures, comparing to the observers' measures.

The same trend is shown in the scatter plots of the FWHM algorithm compare to the single observers: an underestimation is shown, but it is less pronounced than the one showed by the flex points.

Bland-Altman plots follow.

The Bland-Altman plots regarding the comparison between observers (Figure 6.10), show the presence of some outliers. These are the same in all the plots; it means that the vessel section in question could present a border not well marked, then every observer considered the border in different way.

The Bland-Altman plots regarding the software against the single observers (Figure 6.12 and Figure 6.11) present outliers in the same way, but these are due to problems showed by the model to fit some vessels.

The last step is a more depth statistical analysis.

Hypothesis: There are no significant differences among the 5 subjects (Paola, Manuel, Peter, FWHM and Flex), each having 3 measurements.

When the results are $p \leq 0.05$, the hypothesis will be reject, i.e. significant differences found.

In the present chapter I show the results regarding all 60 vessels together, and in Appendix I relate the results concerning arteries and veins taken distinctly.

The Table A.1 in Appendix contains some descriptive parameters (mean, standard error and 95% interval of confidence) regarding all the measures taken to generate the ground truth; the Tables A.2 and A.3 in Appendix regarding the same measures subdivided into arteries and veins.

The next step was to realize a mean of the three measurements taken by each observer, in order to obtain 5 total measures: Paola (mean of Paola1, Paola2 and Paola3), Manuel (mean of Manuel1, Manuel2 and Manuel3), Peter (mean of Peter1, Peter2 and Peter3), FWHM (mean of FWHM1, FWHM2 and FWHM3) and Flex (mean of Flex1, Flex2 and Flex3); first a descriptive statistics was done like previously, Table 6.1, and then the significance values and the 95% Confidence Interval of the difference have been calculated, Table A.7 (Appendix) considering all the vessels together.

In the present chapter I report the results obtained for arteries and veins taken separately: Table 6.2 and Table 6.3 .

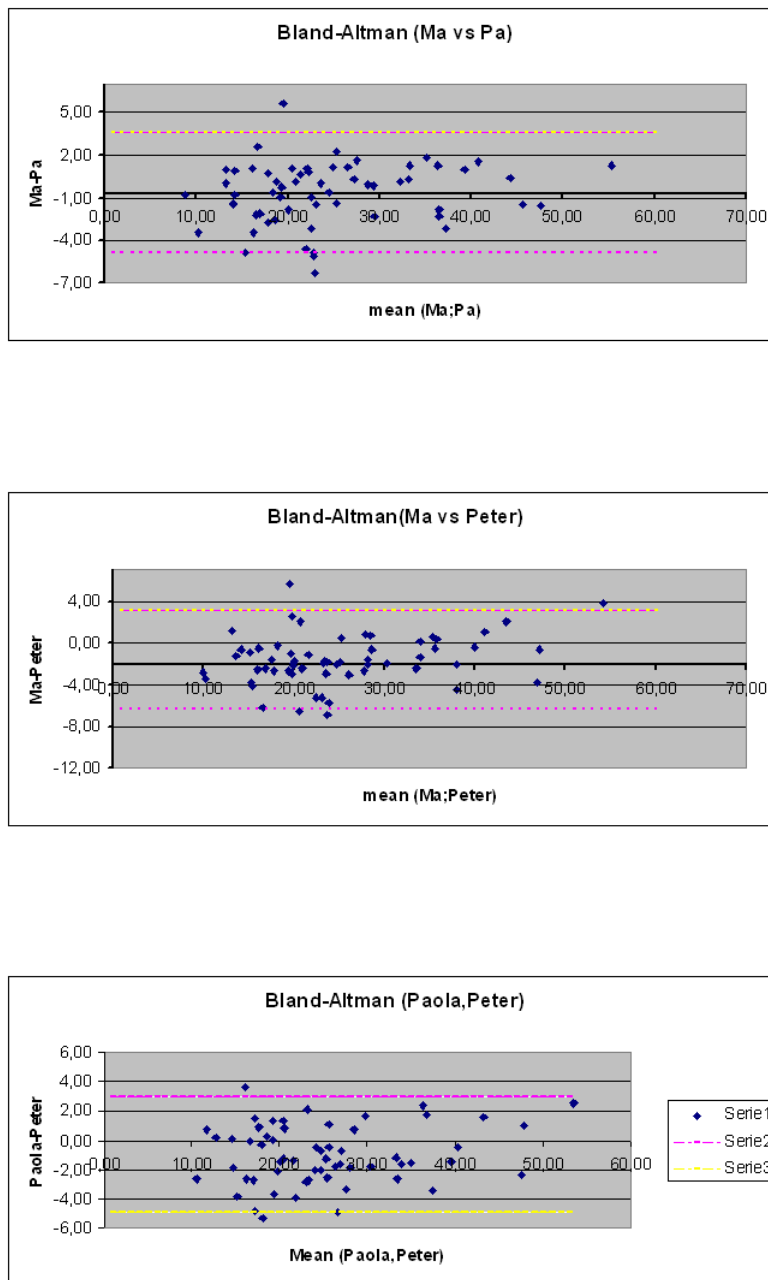


Figure 6.10: Bland Altman, comparison between the three observers

Finally, a statistical comparison between Man (obtained as mean of Paola, Manuel and Peter) and Machine. The general results are shown in Table A.8 (Appendix) and the subdivided results in Table 6.4 and Table 6.5.

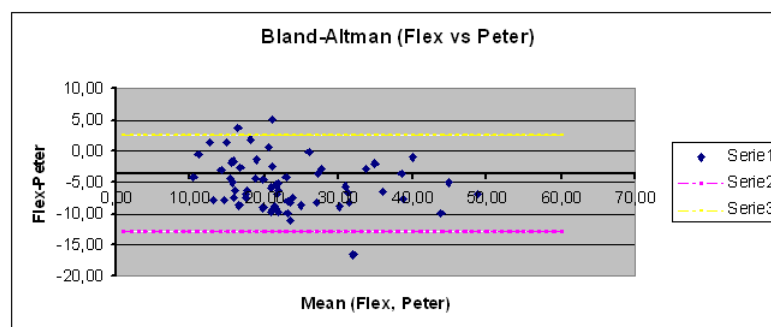
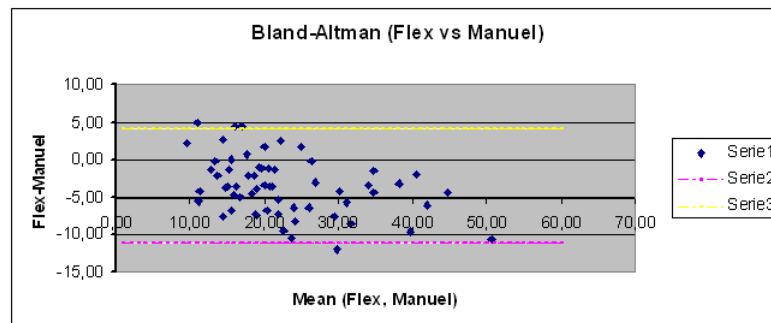
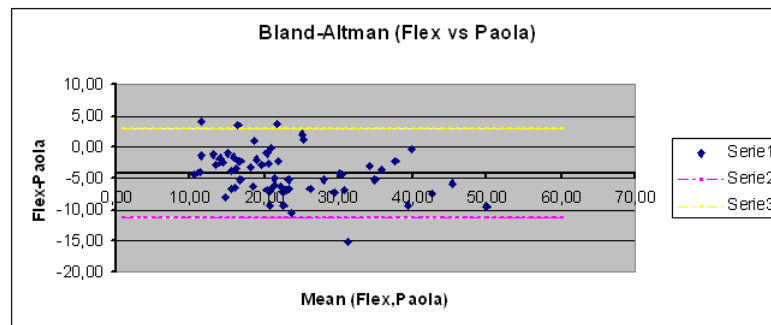


Figure 6.11: Bland Altman, comparison between the Flex points technique and the observers

The paired t-test emphasizes that there are significant differences. There are significant differences inter observers and also between the single observers and the software.

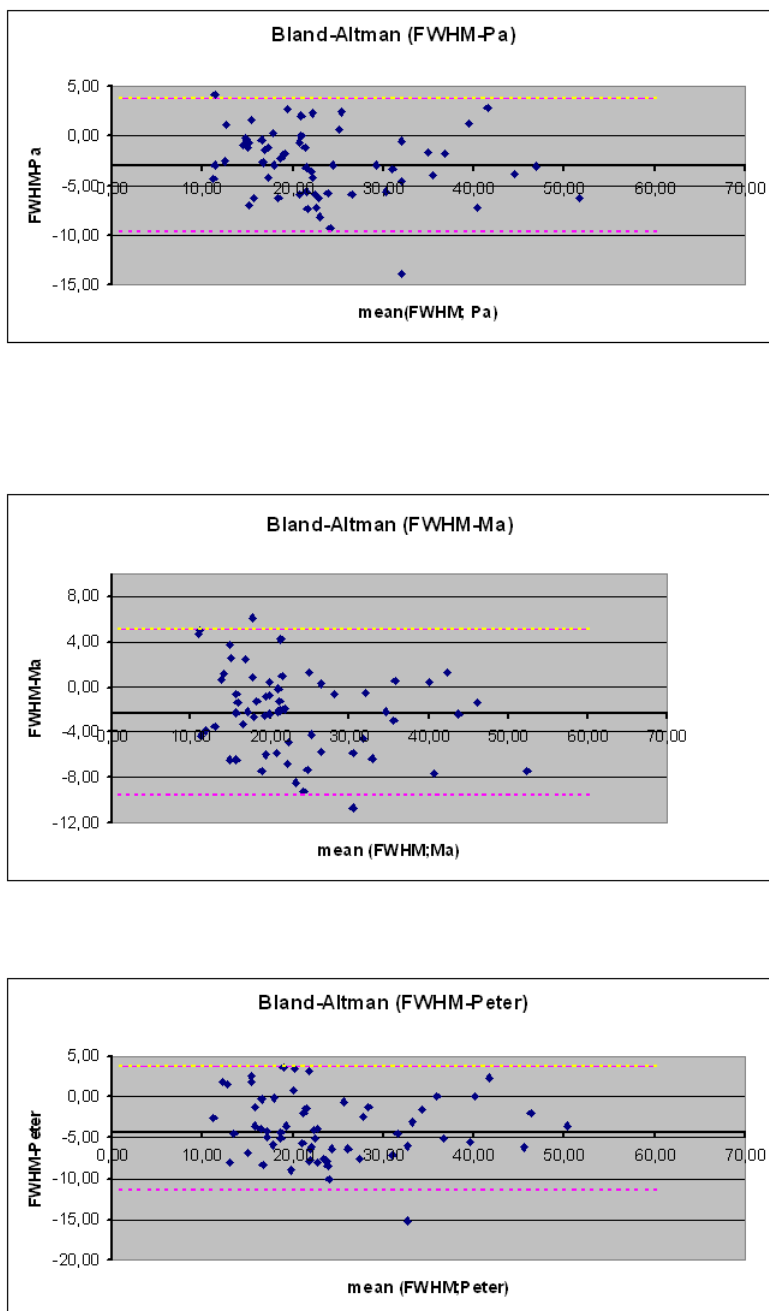
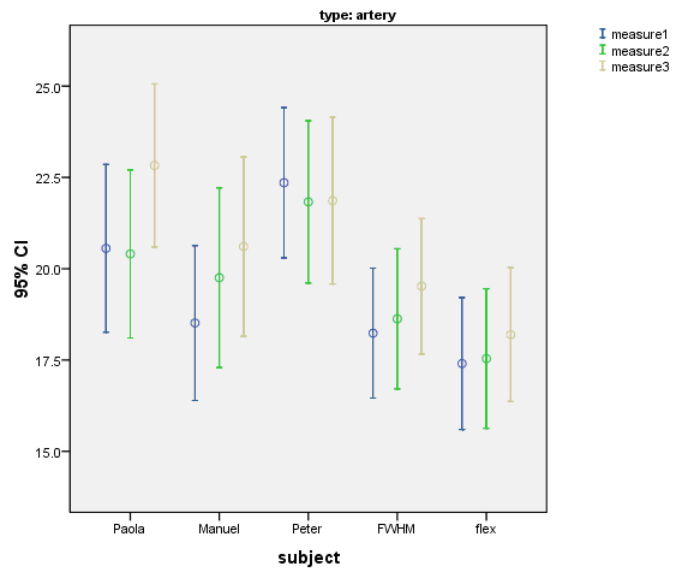
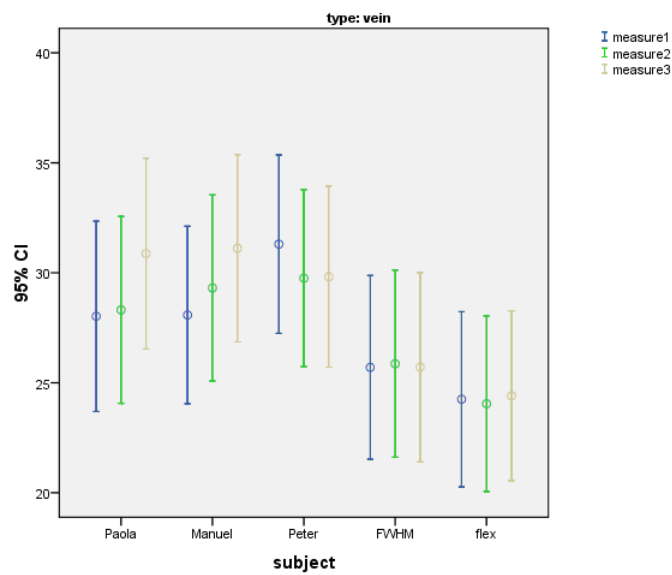


Figure 6.12: Bland Altman, comparison between the FWHM algorithm and the observers

Looking at the mean differences we observe that the software (both flex points technique and FWHM) is prone to underestimate the measures, comparing to the observers' mea-



(a) Arteries' measures



(b) Veins' measures

Figure 6.13: Graphical representation of all measures, respective means and confidence intervals are reported

asures; to be more precise the flex points technique is the one that under estimate more. This is shown in Figure 6.13, regarding all the measures taken, means and confidence intervals are reported.

This means that the differences between the means of the subjects is significant

	mean	std.err	95% confidence interval
Paola	25.04	1.25	22.53/27.54
Manuel	24.40	1.30	21.81/26.99
Peter	26.02	1.21	23.59/28.45
FWHM	22.16	1.18	19.91/24.51
Flex	20.87	1.10	18.66/23.07

Table 6.1: Descriptive Statistics of all the measures

and it is not due to chance. Therefore we can not assert that our software returns indiscernible results from those human.

Due to this conclusion we decided to modify the software, therefore a new statistical analysis follows.

6.6 Conclusions

In the first part of the present chapter I have demonstrated the stability and the solidity of the Extended MHM.

Afterwards, I explained how we tackle the validation. We generated a ground truth with fifteen measures for each vessel: nine were taken by the observers (three measures each observer) and six were taken by the software with the two different algorithms used, flex points technique and FWHM.

I described briefly the statistics that we have used for validating the model.

Finally, in the last part of the chapter, I reported the results outcome. The statistics results point out that the flex points technique underestimates the widths more than FWHM.

Due to this reason, we decided to improve the FWHM algorithm and drop the flex points technique to measure the vessel's width.

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Paola	<u>Manuel</u>	1.64	0.37	0.000	0.89/2.38
	<u>Peter</u>	-0.75	0.37	0.048	-1.49/-0.02
	<u>FWHM</u>	2.47	0.59	0.000	1.28/3.65
	<u>Flex</u>	3.55	0.62	0.000	2.32/4.78
Manuel	<u>Paola</u>	-1.64	0.37	0.000	-2.38/-0.89
	<u>Peter</u>	-2.39	0.41	0.000	-3.20/-1.58
	<u>FWHM</u>	0.83	0.67	0.000	0.51/2.17
	<u>Flex</u>	1.91	0.68	0.000	0.56/3.27
Peter	<u>Paola</u>	0.75	0.37	0.048	0.02/1.49
	<u>Manuel</u>	2.39	0.41	0.000	1.58/3.20
	<u>FWHM</u>	3.22	0.68	0.000	1.86/4.58
	<u>Flex</u>	4.30	0.68	0.000	2.94/5.67
FWHM	<u>Paola</u>	-2.47	0.59	0.000	-3.65/-1.28
	<u>Manuel</u>	-0.83	0.67	0.000	-2.17/-0.51
	<u>Peter</u>	-3.22	0.68	0.000	-4.58/-1.86
	<u>Flex</u>	1.08	0.20	0.000	0.67/1.49
Flex	<u>Paola</u>	-3.55	0.62	0.000	-4.78/-2.32
	<u>Manuel</u>	-1.91	0.68	0.000	-3.27/-0.56
	<u>Peter</u>	-4.30	0.68	0.000	-5.67/-2.94
	<u>FWHM</u>	-1.08	0.20	0.000	-1.49/-0.67

Table 6.2: Pairwise Comparisons for arteries

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Paola	Manuel	-0.43	0.32	0.193	-1.08/0.22
	<u>Peter</u>	-1.22	0.36	0.002	-1.95/ -0.49
	<u>FWHM</u>	3.31	0.68	0.000	1.95/4.68
	<u>Flex</u>	4.84	0.71	0.000	3.41/6.26
Manuel	Paola	0.43	0.32	0.193	-0.22/1.08
	Peter	-0.79	0.42	0.073	-1.64/0.06
	<u>FWHM</u>	3.75	0.60	0.000	2.55/4.94
	<u>Flex</u>	5.27	0.60	0.000	4.06/6.47
Peter	<u>Paola</u>	1.22	0.36	0.002	0.49/1.95
	Manuel	0.79	0.42	0.073	-0.06/1.64
	<u>FWHM</u>	4.54	0.73	0.000	3.07/6.00
	<u>Flex</u>	6.06	0.73	0.000	4.59/7.53
FWHM	<u>Paola</u>	-3.31	0.68	0.000	-4.68/-1.95
	<u>Manuel</u>	-3.75	0.60	0.000	-4.94/-2.55
	<u>Peter</u>	-4.54	0.73	0.000	-6.00/-3.07
	<u>Flex</u>	1.52	0.25	0.000	1.02/2.02
Flex	<u>Paola</u>	-4.84	0.71	0.000	-6.26/-3.41
	<u>Manuel</u>	-5.27	0.60	0.000	-6.47/-4.06
	<u>Peter</u>	-6.06	0.73	0.000	-7.53/-4.59
	<u>FWHM</u>	-1.52	0.25	0.000	-2.02/-1.02

Table 6.3: Pairwise Comparisons for veins

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Man	<u>FWHM</u>	2.17	0.60	0.001	0.98/3.37
	<u>Flex</u>	3.26	0.62	0.000	2.01/4.50
FWHM	<u>Man</u>	-2.17	0.60	0.001	-3.37/-0.98
	<u>Flex</u>	1.08	0.20	0.000	0.68/1.48
Flex	<u>Man</u>	-3.26	0.62	0.000	-4.50/-2.01
	<u>FWHM</u>	-1.08	0.20	0.000	-1.48/-0.68

Table 6.4: Pairwise Comparison between Man and Machines, arteries

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Man	<u>FWHM</u>	3.86	0.64	0.000	2.59/5.14
	<u>Flex</u>	5.39	0.65	0.000	4.09/6.69
FWHM	<u>Man</u>	-3.86	0.64	0.000	-5.14/-2.59
	<u>Flex</u>	1.52	0.25	0.000	1.02/2.02
Flex	<u>Man</u>	-5.39	0.65	0.000	-6.69/-4.09
	<u>FWHM</u>	-1.52	0.25	0.000	-2.02/-1.02

Table 6.5: Pairwise Comparison between Man and Machines, veins

Chapter 7

Optimization of the algorithm to measure the width

7.1 Introduction

In this chapter I will explain how I modify the FWHM algorithm in order to solve the problem of the width underestimation. Once completed, a new statistical analysis is conducted and a discussion of the new outcome results occurs.

7.2 Optimization results

Analyzing carefully the means of the differences, we have seen, as mentioned in previous chapter, that our software underestimates the width of the vessels compared to the observers. The underestimation is bigger if we use the flex points technique.

Due to this reason, we decided to improve the FWHM algorithm and not to use the flex points technique.

The FWHM algorithm takes the width of the vessels as the distance between the two mid points between the maximum and minimum values of the fit, but this returns an underestimation of the widths. Therefore we chose to increase the points in which we take the width, then we will add to the minimum values of the fit something more than half height.

To find this value we proceeded with an optimization technique.

The FWHM (Full Width Half Maximum) works by adding to the minimum values of the fit half of the distance between maximum and minimum values, as previously explained. We noticed that this half height was not enough, because the results were

underestimated.

We decided to optimize the algorithm by using an optimization technique: the minimization of the error function.

We will call δ the constant value that multiplies the height calculated between the minimum and maximum values of the fitness, in the FWHM δ is equal to 0.5.

The first step has been to calculate the error function of FWHM algorithm compared to the single observers first, and the observers considered like a single person (Man is given by the mean of Paola, Manuel and Peter) finally.

The error function is given by the Equation 7.1:

$$e(\delta) = \sqrt{\sum_{k=1}^N (w_{p_k}(\delta) - w_{o_k}(\delta))^2} \quad (7.1)$$

where N = number of vessels.

The Equation 7.1 consists in calculating the difference of width for every vessel (calculated with the FWHM and measured by the observers), to square these difference, sum the values and calculate the square root. Therefore the function error will be a number.

We will consider different values for $\delta = [0.550.60.650.7]$, in order to find the value of δ that returns the minimum value for the error function:

$$\delta^* = \min_{\delta}(e(\delta)) \quad (7.2)$$

δ^* will be the value to multiple the height which allows to obtain the more similar width of the cross section, as compared to the ones taken by the observers.

We run the software with the δ values chosen, some fitness are shown below, Figure 7.1:

The values of the error function for each δ value are shown in Table 7.1:

The Table 7.1 shows that the best value for δ is 0.6 because, in correspondance to this value there are the minimum values for the error function.

For this reason we call the algorithm FW60M (Full Width Sixty Maximum), because we take the 60% of the height between the maximum and minimum as the points to measure the width of the vessel (along the x-axis).

7.3 Statistical analysis for FW60M

Thus a new statistical analysis is required, we proceede the same way as above. The Scatter plot of the FW60M algorithm compared with the observers, Figure A.1 (Ap-

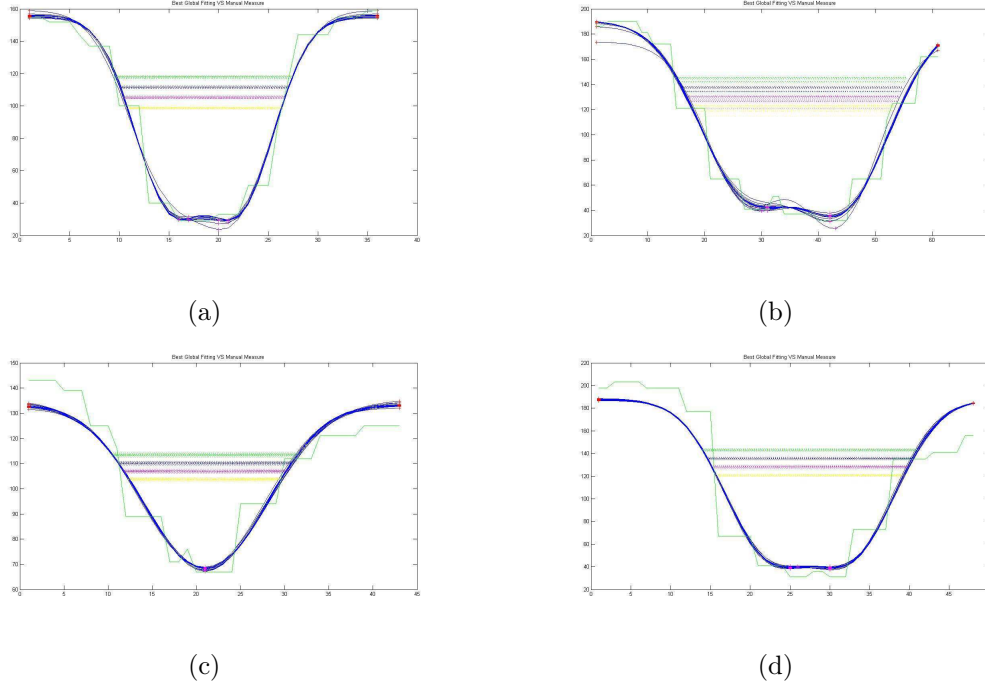


Figure 7.1: Validation: fitness showing the width calculated at 55%, 60%, 65% and 70% of the height

δ value	error function FWM(δ)-Pa	error function FWM(δ)-Ma	error function FWM(δ)-Pt	error function FWM(δ)-Obs
0.50	34.66	33.48	42.07	63.97
0.55	29.7	32.11	37.19	57.55
<u>0.60</u>	<u>28.34</u>	<u>32.3</u>	<u>34.58</u>	<u>55.15</u>
0.65	30.06	35.39	34.62	57.91
0.70	34.16	40.48	36.71	64.44

Table 7.1: Values of the Error function for each δ

pendix), shows now a major agreement between the software and the observers: it is evident that there is no more underestimations and the measures now are well distributed along the trendline.

The Bland-Altman plots keep to show some outliers (see Figure A.2, Appendix), as previously, these are due to inaccuracies shows by the model to fit some vessels.

The new statistical analysis follows, as in previous chapter: the values regarding all the vessels considered together are reported in Appendix, and in the present chapter the values concerning arteries and veins are separately listed.

	mean	std.err	95% confidence interval
Paola	25.04	1.25	22.53/27.54
Manuel	24.40	1.30	21.81/26.99
Peter	26.02	1.21	23.59/28.45
FW60M	24.72	1.28	22.16/27.28

Table 7.2: Descriptive Statistics of all measures

Looking at Table 7.2 we can observe that the mean of the FW60M algorithm is higher than the FWHM's ones. This mean is more similar to the means of the observers.

Observing Table 7.3, relevant to the arteries, we observe that there are no significant differences between the FW60M algorithm and Paola and Peter, instead there is a significant difference with Manuel ($p\text{Value} = 0.025 \leq 0.05$).

We analyze with more attention this value: the mean difference between FW60M and Manuel is 1.841 pixel and the respective pValue is 0.025; but the mean difference between Manuel and Paola is 1.639 (comparable with 1.841) and its respective pValue is 0.000 (≤ 0.025). This allows to conclude that the difference between Manuel and Paola is statistically more significant (also if smaller in order of pixels) than the difference between Manuel and FW60M.

Therefore Manuel is more similar to FW60M than Paola, relevant to arteries.

Looking at table 7.4, regarding the veins, we observe that there is a significant difference between FW60M and Peter. Table A.9, which considers all the vessels together without distinction in artery or vein, allows to conclude that there are significant differences between the observers; whereas concerning FW60M the only significant difference is with Peter. As previously, if we compare the means with the pValues we observe that the mean difference between FW60M and Peter is -1.30 pixel with pValue equal

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Paola	<u>Manuel</u>	1.64	0.37	0.000	0.89/2.38
	<u>Peter</u>	-0.75	0.37	0.048	-1.49/-0.02
	FW60M	-0.20	0.63	0.748	-1.45/1.05
Manuel	<u>Paola</u>	-1.64	0.37	0.000	-2.38/-0.89
	<u>Peter</u>	-2.39	0.41	0.000	-3.20/-1.58
	<u>FW60M</u>	-1.84	0.78	0.024	-3.40/-0.29
Peter	<u>Paola</u>	0.75	0.37	0.048	0.02/1.49
	<u>Manuel</u>	2.39	0.41	0.000	1.58/3.20
	FW60M	0.55	0.74	0.461	-0.93/2.03
FW60M	Paola	0.20	0.63	0.748	-1.05/1.45
	<u>Manuel</u>	1.84	0.78	0.024	0.29/3.40
	Peter	-0.55	0.74	0.461	-2.03/0.93

Table 7.3: Pairwise Comparisons for arteries

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Paola	Manuel	-0.43	0.32	0.193	-1.08/0.22
	<u>Peter</u>	-1.22	0.36	0.002	-1.95/ -0.49
	FW60M	0.87	0.75	0.253	-0.62/2.36
Manuel	Paola	0.43	0.32	0.193	-0.22/1.08
	Peter	-0.79	0.42	0.073	-1.64/0.06
	FW60M	1.30	0.68	0.067	-0.06/2.36
Peter	<u>Paola</u>	1.22	0.36	0.002	0.49/1.95
	Manuel	0.79	0.42	0.073	-0.06/1.64
	<u>FW60M</u>	2.09	0.82	0.016	0.45/3.73
FW60M	Paola	-0.87	0.75	0.253	-2.36/0.62
	Manuel	-1.30	0.68	0.067	-2.67/0.06
	<u>Peter</u>	-2.09	0.82	0.016	-3.73/-0.45

Table 7.4: Pairwise Comparisons for veins

to 0.0288; otherwise the mean difference between Peter and Manuel is 1.62 pixel and the respective pValue is 0.0000. Therefore Peter is statistically more similar to FW60M than Manuel.

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Man	FW60M	-0.50	0.68	0.471	-1.86/0.87

Table 7.5: Pairwise Comparison between Man and Machine (FW60M), arteries

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Man	FW60M	1.42	0.72	0.058	-0.02/2.86

Table 7.6: Pairwise Comparison between Man and Machine (FW60M), veins

As last analysis, we have compared the FW60M algorithm with all the observers together (considered like an unique observer, it is provided as a mean of Paola, Manuel and Peter).

Tables 7.5, 7.6 and table A.10 (Appendix) show that the mean differences are not significant.

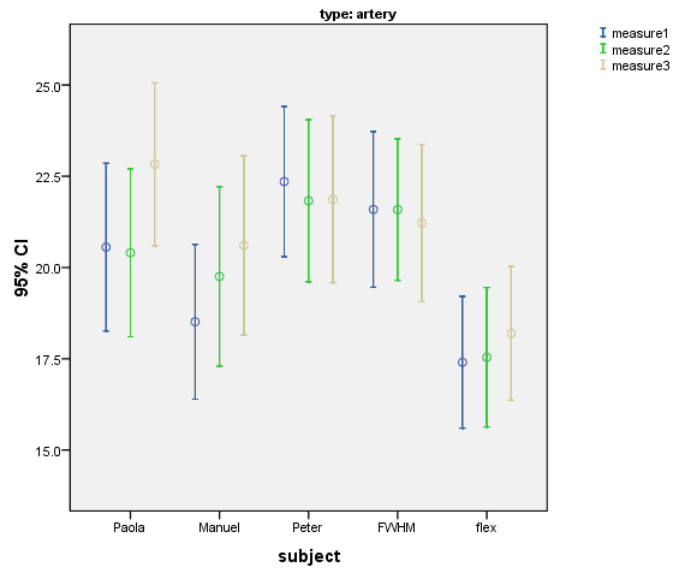
The total mean difference, considering all the vessels together, is very low: our software underestimates 0.43 pixel, the underestimation is 0.50 pixel for the arteries and -1.42 pixel for the veins.

Finally our model, with the appropriate alteration done to the FWHM algorithm, is indiscernible from the observers, and they can be used interchangeably.

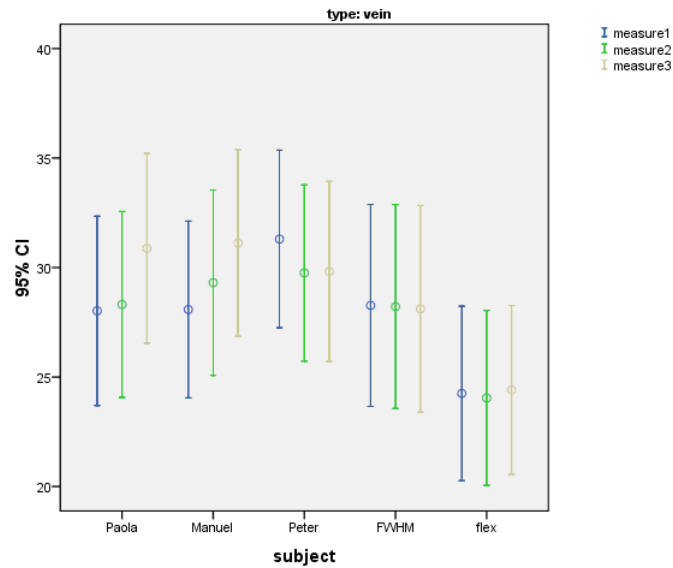
7.4 Conclusions

In the present chapter, I modified the FWHM algorithm into FW60M. I used the minimization of the error function to find the percentage of the height, that allows to obtain widths comparable with the ones taken by the observers.

A statistical analysis with the new algorithm, showed that our model, extended MHM and FW60M, is indiscernible from the observers. Therefore it can replace the humans during the measurement phase.



(a) Arteries' measures



(b) Veins' measures

Figure 7.2: Graphical representation of all measures, respective means and confidence intervals are reported

Chapter 8

Conclusions and future work

8.1 Conclusions

The work described in this thesis is relevant to an algorithm to automatically extract width measures from retinal vessels.

This plays a critical part in the structured analysis of the retina, and it is potentially useful in the automated diagnosis of eye diseases such as diabetic retinopathy.

The first part of my work concerns the extension of an Multiresolution Hermite model to fit the intensity profile of a retinal vessel. My goal was to obtain a model that could fit a broad spectrum of intensity retinal vessel profiles: non symmetric profiles, symmetric profiles and profile with or without central reflex.

The fitness that are obtained show that the model can follow almost perfectly all the vessels being analyzed. Furthermore, it demonstrated to be accurate and stable due to the fact that it can fit the same cross section for fifteen times without major alterations. The model presented some problems just with three vessels out of sixty in total: it had problems to fit univocally these vessels. These problems are due to the low contrast of the image between the background and the vessel.

Some stability issues were also seen in few vessels (no more than ten out of 60): looking at the fifteen runs we noticed that about $4/5$ times out of 15, the fitness do not follow the correct fit.

Anyway these issues represent the minority of the cases; hence we can conclude that the Extended Multiresolution Hermite model is stable and accurate.

In the second part of my thesis, we considered an algorithm to measure the width of the cross-section previously selected .

The FWHM demonstrated to underestimate the width of the vessel, compared with the

one taken by the observers.

Therefore we modified the FWHM in FW60M: we take as the 2 points to measure the distance, the 60% of the height between the minimum and the maximum values on the left, and the minimum and maximum values on the right.

The statistical analysis performed shows that the measures of width given by the software do not have significant differences from those taken by the observers.

The paired t-test used shows also that there are significant differences inside the three observers. These differences are statistically different between Paola and Manuel against Peter: Peter is inclined to overestimate all the widths compared to Paola's and Manuel's measures.

Another important aspect to consider is that FW60M demonstrated to be more consistent during the three sessions of measurement than the observers. Its mean is considered not to change during the three measurements; whereas Paola's mean and Manuel's one are about 2 pixels more in the third measure than in the first one.

The validation process that we used for this work is different from the validation in which there is calibrated instrument to provide exact measures (e.g. industrial inspection).

In this work, the observers are the instruments, then it is not possible to know how much our measures are accurate; there is not an absolute truth to compare with the results provided by the software. Therefore the model is considered efficient if it is indistinguishable from one of the observer, and if it is accurate as the observers.

The results demonstrate that our model is accurate and indistinguishable from the observers.

Then the Extended Multiresolutional Hermite model could be used in practice to measure the width of retinal vessels instead of humans.

8.2 Future work

Although the results presented in my thesis shown the effectiveness of our approach, it could be developed in a number of ways.

We did not develop this aspect due to lack of time. This may be subject of further analysis in the future.

As previously mentioned, the model had some problems to fit some vessels. These vessels were characterized with low contrast between the background and the vessel, then the border of the section was not well defined.

Future work may improve the model in order to fit perfectly every kind of vessels;

specially the vessels that are contained in fundus images with low contrast.

Another aspect to be improved is the way in which the software returns the final width of the vessel. We used an histogram, reporting all the fifteen widths returned by the software, and we chose the peak as the final width. Sometimes identification issues occurred: the histogram could have more than an unique peak, or it could be difficult to identify the exact value. An improvement of this technique is then required.

With regard to the validation phase, the results showed that there are significant differences between Peter and the other two observers (Paola and Manuel): Peter was included to overestimate the widths. In the future an uniformity in the measures has to be guaranteed: a meeting with all the observers has to support the paper protocol.

Finally, a major ground truth needs to be generated. Our validation phase concerns sixty vessels, but increasing this number will provide more accurate results.

Appendix A

Statistics

A.1 Descriptive statistics: Tables

	mean	std.err	95% confidence interval
Paola1	24.17	1.26	21.64/26.69
Paola2	24.23	1.26	21.71/26.74
Paola3	26.72	1.27	24.18/29.26
Manuel1	23.14	1.25	20.64/25.64
Manuel2	24.37	1.32	21.73/27.02
Manuel3	25.69	1.36	22.97/28.40
Peter1	26.68	1.23	24.23/29.13
Peter2	25.66	1.21	23.24/28.08
Peter3	25.71	1.23	25.25/28.18
FWHM1	21.84	1.18	19.48/24.20
FWHM2	22.13	1.20	19.72/24.53
FWHM3	22.51	1.18	20.5/24.87
Flex1	20.71	1.13	18.46/22.97
Flex2	20.68	1.13	18.42/22.95
Flex3	21.20	1.09	19.02/23.38

Table A.1: Descriptive Statistics for all vessels and measures

	mean	std.err	95% confidence interval
Paola1	20.56	1.13	18.31/22.81
Paola2	20.40	1.13	18.15/22.66
Paola3	22.83	1.09	20.64/25.01
Manuel1	18.51	1.04	16.44/20.59
Manuel2	19.76	1.21	17.35/22.17
Manuel3	20.61	1.20	18.20/23.01
Peter1	22.35	1.01	20.34/24.37
Peter2	21.83	1.09	19.65/24.01
Peter3	21.87	1.12	19.63/24.10
FWHM1	18.24	0.87	16.49/19.98
FWHM2	18.63	0.94	16.75/20.51
FWHM3	19.52	0.91	17.70/21.34
Flex1	17.40	0.88	15.63/19.17
Flex2	17.54	0.93	15.67/19.41
Flex3	18.20	0.90	16.40/19.99

Table A.2: Descriptive Statistics of all the measures for arteries

	mean	std.err	95% confidence interval
Paola1	28.02	2.11	23.80/32.24
Paola2	28.31	2.07	24.17/32.46
Paola3	30.88	2.12	26.64/35.11
Manuel1	28.08	1.97	24.14/32.02
Manuel2	29.31	2.07	25.18/33.44
Manuel3	31.12	2.08	26.96/35.27
Peter1	31.30	1.98	27.34/35.26
Peter2	29.75	1.96	25.83/33.68
Peter3	29.82	2.01	25.81/33.84
FWHM1	25.70	2.04	21.62/29.78
FWHM2	25.87	2.07	21.72/30.01
FWHM3	25.71	2.10	21.51/29.91
Flex1	24.25	1.95	20.36/28.14
Flex2	24.04	1.95	20.14/27.94
Flex3	24.41	1.89	20.64/28.18

Table A.3: Descriptive Statistics of all the measures for veins

	mean	std.err	95% confidence interval
Paola1	24.17	1.26	21.64/26.69
Paola2	24.23	1.26	21.71/26.74
Paola3	26.72	1.27	24.18/29.26
Manuel1	23.14	1.25	20.64/25.64
Manuel2	24.37	1.32	21.73/27.02
Manuel3	25.69	1.36	22.97/28.40
Peter1	26.68	1.23	24.23/29.13
Peter2	25.66	1.21	23.24/28.08
Peter3	25.71	1.23	25.25/28.18
FW60M1	24.82	1.28	22.26/27.38
FW60M2	24.79	1.27	22.26/27.33
FW60M3	24.55	1.31	21.93/27.17

Table A.4: Descriptive Statistics for all vessels and measures

	mean	std.err	95% confidence interval
Paola1	20.56	1.13	18.31/22.81
Paola2	20.40	1.13	18.15/22.66
Paola3	22.83	1.09	20.64/25.01
Manuel1	18.51	1.04	16.44/20.59
Manuel2	19.76	1.21	17.35/22.17
Manuel3	20.61	1.20	18.20/23.01
Peter1	22.35	1.01	20.34/24.37
Peter2	21.83	1.09	19.65/24.01
Peter3	21.87	1.12	19.63/24.10
FW60M1	21.59	1.04	19.51/23.68
FW60M2	21.59	0.95	19.69/23.49
FW60M3	21.22	1.06	19.10/23.33

Table A.5: Descriptive Statistics of all the measures for arteries

	mean	std.err	95% confidence interval
Paola1	28.02	2.11	23.80/32.24
Paola2	28.31	2.07	24.17/32.46
Paola3	30.88	2.12	26.64/35.11
Manuel1	28.08	1.97	24.14/32.02
Manuel2	29.31	2.07	25.18/33.44
Manuel3	31.12	2.08	26.96/35.27
Peter1	31.30	1.98	27.34/35.26
Peter2	29.75	1.96	25.83/33.68
Peter3	29.82	2.01	25.81/33.84
FW60M1	28.27	2.25	23.77/32.78
FW60M2	28.22	2.27	23.67/32.76
FW60M3	28.11	2.30	23.50/32.72

Table A.6: Descriptive Statistics of all the measures for veins

A.2 Statistical Analysis: paired t-test, Tables

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Paola	<u>Manuel</u>	0.64	0.28	0.026	0.08/1.20
	<u>Peter</u>	-0.98	0.26	0.000	-1.50/ -0.46
	<u>FWHM</u>	2.88	0.45	0.000	1.98/3.77
	<u>Flex</u>	4.17	0.47	0.000	3.23/5.12
Manuel	<u>Paola</u>	-0.64	0.28	0.026	-1.20/ -0.08
	<u>Peter</u>	-1.62	0.31	0.000	-2.23/ -1.00
	<u>FWHM</u>	2.24	0.48	0.000	1.28/3.20
	<u>Flex</u>	3.53	0.50	0.000	2.53/4.54
Peter	<u>Paola</u>	0.98	0.26	0.000	0.46/1.50
	<u>Manuel</u>	1.62	0.31	0.000	1.00/2.23
	<u>FWHM</u>	3.86	0.50	0.000	2.86/4.85
	<u>Flex</u>	5.15	0.51	0.000	4.13/6.17
FWHM	<u>Paola</u>	-2.88	0.45	0.000	-3.77/ -1.98
	<u>Manuel</u>	-2.24	0.48	0.000	-3.20/ -1.28
	<u>Peter</u>	-3.86	0.50	0.000	-4.85/ -2.86
	<u>Flex</u>	1.29	0.16	0.000	0.97/1.62
Flex	<u>Paola</u>	-4.17	0.47	0.000	-5.12/ -3.23
	<u>Manuel</u>	-3.53	0.50	0.000	-4.54/ -2.53
	<u>Peter</u>	-5.15	0.51	0.000	-6.17/ -4.13
	<u>FWHM</u>	-1.29	0.16	0.000	-1.62/ -0.97

Table A.7: Pairwise Comparisons for all the measures

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Man	<u>FWHM</u>	2.99	0.45	0.000	2.10/3.88
	<u>Flex</u>	4.29	0.47	0.000	3.35/5.22
FWHM	<u>Man</u>	-2.99	0.45	0.000	-3.88/ -2.10
	<u>Flex</u>	1.29	0.16	0.000	0.97/1.62
Flex	<u>Man</u>	-4.29	0.47	0.000	-5.22/ -3.35
	<u>FWHM</u>	-1.29	0.16	0.000	-1.62/ -0.97

Table A.8: Pairwise Comparison between Man and Machines, all vessels

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Paola	<u>Manuel</u>	0.64	0.28	0.026	0.08/1.20
	<u>Peter</u>	-0.98	0.26	0.000	-1.50/ -0.46
	FW60M	0.32	0.49	0.5180	-0.66/ -1.29
Manuel	<u>Paola</u>	-0.64	0.28	0.026	-1.20/ -0.08
	<u>Peter</u>	-1.62	0.31	0.000	-2.23/ -1.00
	FW60M	-0.32	0.55	0.5636	-1.43/0.79
Peter	<u>Paola</u>	0.98	0.26	0.000	0.46/1.50
	<u>Manuel</u>	1.62	0.31	0.000	1.00/2.23
	<u>FW60M</u>	1.30	0.55	0.0228	0.19/2.40
FW60M	Paola	-0.32	0.49	0.5180	-1.29/0.66
	Manuel	0.32	0.55	0.5636	-0.79/1.43
	<u>Peter</u>	-1.30	0.55	0.0228	-2.40/ -0.19

Table A.9: Pairwise Comparisons for all the measures

subject(i)	subject(j)	mean (i-j)	std.err	Sign.Value	95% CI
Man	FW60M	0.43	0.51	0.399	-0.58/1.44
FW60M	Man	-0.43	0.51	0.399	-1.44/0.58

Table A.10: Pairwise Comparison between Man and Machine (FW60M), all vessels

A.3 Scatter and Bland-Altman plots

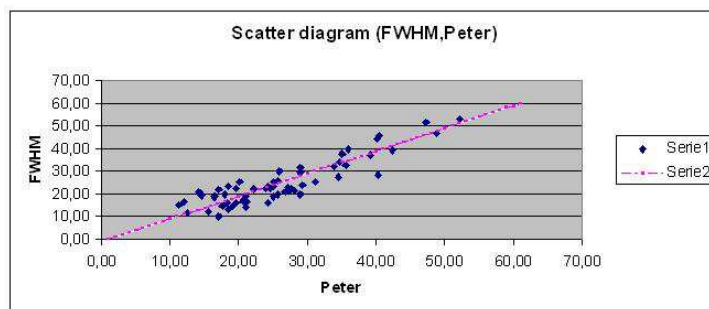
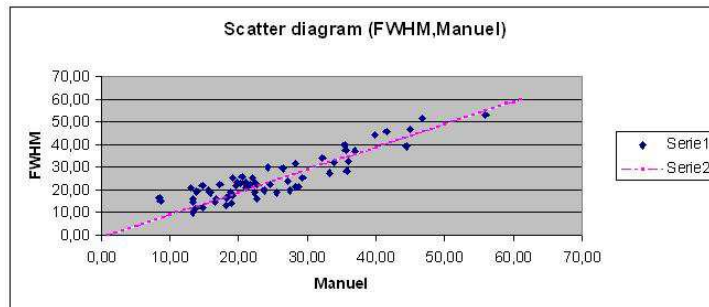
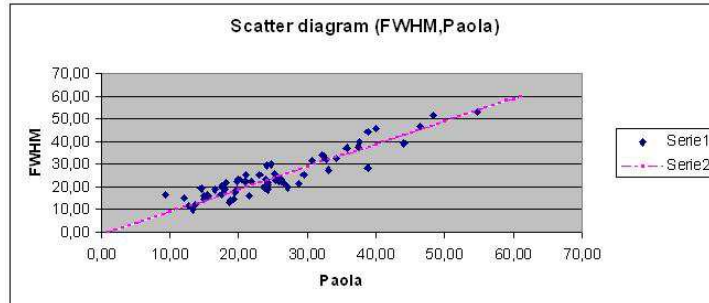


Figure A.1: Scatter plots, comparison between FW60M algorithm and the three observers

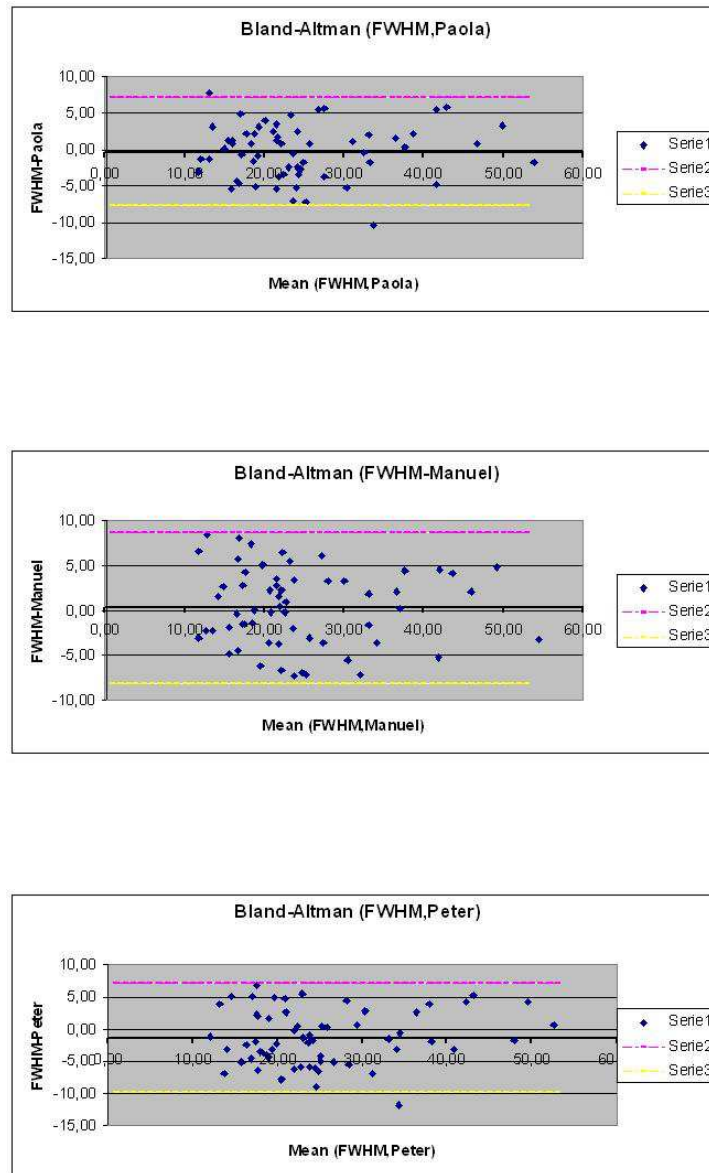


Figure A.2: Bland-Altman plots, comparison between FW60M algorithm and the three observers

Bibliography

- [1] K.H. Fritzsche. Computer vision algorithms for retinal vessel width change detection and quantification. Master's thesis, Rensselaer Polytechnic Institute, Department of Computer Science, 2002.
- [2] N. Patton, T.M. Aslam, T. MacGillivray, I.J. Deary, B. Dhillon, R.H. Eikelboom, K. Yogesan, and I.J. Constable. Retinal images analysis: Concepts, applications and potential. *Progress in Retinal and Eye Research*, 25:99–127, 2006.
- [3] Bryan W. Jones. Webvision; the organization of the retina and visual system. <http://webvision.med.utah.edu/sretina.html>.
- [4] U.S. Department of health and human services - National Institutes of Health. *Diabetic retinopathy - What you should know*. <http://www.nei.nih.gov/health/diabetic/diabeticretino.pdf>.
- [5] J. Cai and M. Bulton. The pathogenesis of diabetic retinopathy: old concepts and new questions. *Eye*, 16:242–260, May 2002.
- [6] H. Narasimha-Iyer, V. Mahadevan, J.M. Beach, and B. Roysam. Improved Detection of the Central Reflex in Retinal Vessels Using a Generalized Dual-Gaussian Model and Robust Hypothesis Testing. *IEEE Transactions of Information Technology in Biomedicine*, 12, May 2008.
- [7] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R.L. Kennedy. Measurement of Retinal Vessel Widths From Fundus Images Based on 2D-Modeling. *IEEE Transactions on Medical Imaging*, 23(10):1196–1204, October 2004.
- [8] L. Zhou, M. S. Rzeszotarski, L. J. Singerman, and J. M. Chokreff. The Detection and Quantification of Retinopathy Using Digital Angiograms. *IEEE Transaction on Medical Imaging*, 13(4):619–626, December 1994.

- [9] L. Wang, A. Bhalerao, and R. Wilson. Analysis of Retinal Vasculature Using a Multiresolution Hermite Model. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 26(2):137–152, February 2007.
- [10] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, UK, 1970.
- [11] L. Pedersen, M. Grunkin, B. Ersbøll, K. Madsen, M. Larsen, N. Christoffersen, and U. Skands. Quantitative measurement of changes in retinal vessel diameter in ocular fundus images. *Elsevier-Pattern Recognition Letters*, 21(13-14):1215–1223, December 2000.
- [12] O. Brinchmann-Hansen and O. Engvold. Fundus photography of width and intensity profiles of the blood column and the light reflex in retinal vessels. *Acta Ophthalmologica*, Suppl. 179:20–28, 1986.
- [13] S.M.B. Rassam, V. Patel, O. Brinchmann-Hansen, and O. Engvold. Accurate vessel width measurement from fundus photographs: a new concept. *British Journal of Ophthalmology*, 78:24–29, 1994.
- [14] N. Chapman, N. Witt, X. Gao, A. A. Bharath, A. V. Stanton, S. A. Thom, and A. D. Hughes. Computer algorithms for the automated measurement of retinal arteriolar diameters. *Br J Ophthalmol*, 85:74–79, 2001.
- [15] IEEE. *Particle Swarm Optimization: Development, Applications and Resources*, number IV, Perth, Australia, 2001.
- [16] 8th International Mathematica Symposium. *Exploratory Toolkit for Evolutionary and Swarm-based Optimization*, Avignon, June 2006.
- [17] J.M. Bland and D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, pages 307–310, 1986.

Ringraziamenti

Desidero innanzitutto ringraziare coloro che mi hanno seguita in questo periodo di tesi: il professore Alfredo Ruggeri che mi ha dato la possibilità di svolgere il mio lavoro di tesi a Dundee, realizzando uno dei miei più grandi sogni e il professore Emanuele Trucco che, con tanta pazienza, mi ha seguita passo dopo passo in questo progetto ed ha creduto in me anche nei momenti più difficili, spronandomi ad andare avanti e facendomi sentire parte integrante di un gruppo di ricerca.

Un ringraziamento speciale alla mia famiglia che mi ha sempre sostenuta durante il mio percorso universitario e che in quest'anno mi è stata particolarmente vicina, appoggiandomi appieno nella mia decisione di partire, facendo sacrifici per rendere ciò possibile e non facendomi mai sentire sola anche se ero lontana. In particolar modo ringrazio mio papà, che non si è dimenticato mai un giorno di chiamarmi, che mi ha ascoltata lamentarmi per ore quando avevo qualche problema e che mi ha dato la forza di non mollare mai nel mio percorso universitario.

Ringrazio i miei amici del gruppo di studio Esu, i quali hanno reso gli ultimi miei anni di università davvero speciali; mi hanno dato la forza di partire da casa col sorriso ogni mattina all'alba per studiare, perchè sapevo che loro sarebbero stati ad attendermi, per condividere con me momenti di pazzia, dovuti allo studio, momenti di allegria, momenti di confidenze e momenti di vera amicizia. Grazie, siete stati un'ancora di salvezza per me tante volte. Ringrazio in particolare Mattia, Spola e Ciak che nell'ultimo anno sono stati una sorta di fratelli acquisiti, o forse semplicemente veri amici, mi hanno sempre fatto trovare una spalla su cui appoggiarmi se ne avevo bisogno e mi sono stati vicini in modo a dir poco speciale.

Ringrazio le persone che hanno condiviso con me l'ultimo anno da vicino, il Mediterranean Parea: eravamo tutte lontane da casa, ma siamo riuscite a costruirci una sorta di famiglia scozzese con la quale condividere ogni momento, bello e brutto; avete reso quest'ultimo anno semplicemente magnifico e mi avete aiutata a crescere. Ringrazio in particolar modo Steph, che ha avuto la pazienza di aiutarmi ogni volta che la chiamavo, di capire il mio alquanto incerto inglese e di correggere la mia tesi, e Berta, la quale ha

condiviso con me momenti difficili della sua vita, facendomi sentire importante e amata. Grazie ragazze.

Ringrazio Serena, che in quest'anno non passava giorno senza che mi disturbasse in Skype, ma che mi ha fatto sentire la sua vicinanza e la sua amicizia anche se era lontana migliaia di chilometri. Ringrazio Margherita, che c'è sempre stata nella mia vita ogni volta che ne ho avuto bisogno, che mi ha sempre salvata dai mille problemi che combinavo con il mio computer, e che è sempre stata un'amica su cui contare.

E infine ringrazio Maurizio, perchè mi ha supportata in tutti questi anni di università, mi ha ascoltata lamentarmi per gli esami, ha condiviso con me i miei risultati negativi e gioito con me per quelli positivi, perchè ha sempre cercato di tirarmi su quando andavo giù, e perchè è sempre stato al mio fianco anche quando poteva allontanarsi.

Per concludere ringrazio tutte le persone che in generale hanno camminato al mio fianco fino ad ora, condividendo con me un pezzetto della loro vita. Grazie a tutti.

Paola Nicoli