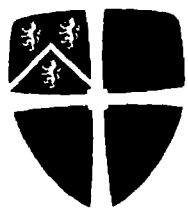
Automated retinal analysis

James . Lowell

Automated Retinal Analysis

by

James A. Lowell



Submitted in conformity with the requirements
for the degree of PhD
Department of Computer Science
University of Durham

Abstract

Automated Retinal Analysis

James A. Lowell

Diabetes is a chronic disease affecting over 2% of the population in the UK [1]. Long-term complications of diabetes can affect many different systems of the body including the retina of the eye. In the retina, diabetes can lead to a disease called diabetic retinopathy, one of the leading causes of blindness in the working population of industrialised countries.

The risk of visual loss from diabetic retinopathy can be reduced if treatment is given at the onset of sight-threatening retinopathy. To detect early indicators of the disease, the UK National Screening Committee have recommended that diabetic patients should receive annual screening by digital colour fundal photography [2]. Manually grading retinal images is a subjective and costly process requiring highly skilled staff.

This thesis describes an automated diagnostic system based on image processing and neural network techniques, which analyses digital fundus images so that early signs of sight threatening retinopathy can be identified. Within retinal analysis this research has concentrated on the development of four algorithms: optic nerve head segmentation, lesion segmentation, image quality assessment and vessel width measurements. This research amalgamated these four algorithms with two existing techniques to form an integrated diagnostic system.

The diagnostic system when used as a 'pre-filtering' tool successfully reduced the number of images requiring human grading by 74.3%: this was achieved by identifying and excluding images without sight threatening maculopathy from manual screening.

Acknowledgements

First and foremost, a special thanks goes to my parents, Anthea and Peter, for their continual support and belief throughout my entire education.

Special thanks goes to Cat, for all her encouragement and total belief throughout my time studying for my PhD.

I would like to thank my academic mentor and principal supervisor Andrew Hunter, for all his patience, encouragement, and guidance. Without which, I may not have got this far.

I wish to thank David Steel, Maged Habib and Ansu Basu for their medical advice and hard work during this research.

Thanks also to Malcolm Munro and Nick Holliman for their thesis advice and proof reading.

Many thanks goes to Barbara (the co-inventor of extreme office ball), for keeping me smiling during the months leading up to submission.

Finally, I would like thank Diabetes UK for funding this research.

Copyright Notice

Declaration

The material contained within this thesis has not previously been submitted for a degree at the University of Durham or any other university. The research reported within this thesis has been conducted by the author unless indicated otherwise. This research has been documented, in part, within the following publications:

- A. Hunter, J. Lowell, J. Owens, and L. Kennedy, *Quantification of Diabetic Retinopathy Using Neural Network and Sensitivity Analysis*,Proceedings of Artificial Neural Networks in Medicine and Biology, 81-86, 2000.
- J. Lowell, A. Hunter, *Automated Detection of Diabetic Retinopathy*, Poster Session, House of Commons, 2001
- A. Hunter, J. Lowell, D. Steel, A. Basu and R. Ryder, *Non-Linear Filtering for Vascular Segmentation and Detection of Venous Beading*,Department of Computer Science, University of Durham, Technical Report, 08/02, 2002.
- J. Lowell, A. Basu, A. Hunter, D. Steel, L. Kennedy, R. Ryder, *A Neural Network System for Diagnosis of Diabetic Retinopathy by White Lesion Detection and Classification*, Diabetes UK, 2002.
- J. Lowell, A. Hunter, *Automated Classification of Referable Diabetic Maculopathy from Fundus Images*, Poster Session, Association for Research in Vision and Ophthalmology, (ARVO), Fort Lauderdale, 2003.
- J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, L. Kennedy, *Optic nerve head segmentation*, IEEE Transactions on Medical Imaging, February, 23(2), 256-64, 2004.

- J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, L. Kennedy, *Measurement of Retinal Vessel Widths From Fundus Images Based on 2-D Modeling*, IEEE Transactions on Medical Imaging, October, 23(10), 1196- 1204, 2004.

- M. Habib, J.Lowell, A. Hunter, D. Steel, *Automated classification of referable diabetic maculopathy from digital fundus images*, North of England Ophthalmological Society. Manchester, March, 2005.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The research contribution of this thesis is the development of techniques for the automated analysis of the retina using digital fundus images. The technical problems associated with retinal analysis have been isolated and have led to the development of four novel algorithms; a deformable model for optic nerve head segmentation; a new region growing algorithm for lesion segmentation based on gradient contrast and compactness; identification of a image quality metric for fundus image quality assessment; and development of a $2D$ vascular segment model for vessel width measurements.

1.1 Introduction

Medical diagnostic imaging has existed since the first decade of the 1900's after the discovery of x-rays. Following World War II and the arrival of the digital computer, new image modalities including ultrasound and magnetic resonance imaging have combined to create an explosion of research into medical image analysis [3].

Diagnostic imaging has enabled non-invasive visualisation of a variety of biological systems, such as the brain, heart, lungs and eyes. Medical diagnosis based on images obtained by ultrasound, computer tomography (CT), magnetic resonance (MR), digital x-rays and digital fundus cameras (see figure 1.2) is now commonplace and has significantly improved the medical care available to patients. With increased medical imaging capabilities, screening programs for the early detection of cancerous tumors or lesions from other diseases such as diabetes have gone nationwide. The

repetitive task of manually assessing large numbers of diagnostic images and the susceptibility of intra-observer variation and error has encouraged the investigation of methods for semi-automated and fully-automated analysis of medical images. Such investigations are usually multidisciplinary due to the complexities of computer vision, and often require implementation of image processing, pattern recognition, and machine learning techniques.

Diabetes mellitus is a chronic disease caused by insufficient insulin being produced by the pancreas or by the ineffectiveness of the insulin produced. In recent years, diabetes mellitus has reached worldwide epidemic proportions. The World Health Organization attributes the increased occurrence of the disease to lifestyle and economic change. In 1985, the worldwide estimate for diabetes was 30 million people; a decade later, 135 million. The WHO's millennium estimation was 177 million worldwide sufferers. This figure is expected to almost double by 2025 [4].

There are two principle forms of diabetes; insulin dependent (type 1) and non insulin dependent (type 2). Insulin dependent diabetes occurs most frequently in children and adolescents and is most commonly caused by the inherited failing of the pancreas to produce enough insulin for the body. Non insulin dependent diabetes accounts for 90% of all cases, and typically effects ageing adults with unhealthy diets, obesity and sedentary lifestyles, and results in the body's inability to process the insulin produced by the pancreas [4]. Increased concentrations of glucose in the blood caused by pancreas deficiencies damage many of the body's systems, but in particular the blood vessels and nerves. Complications of the disease vary with the duration of elevated blood glucose but can include kidney failure, heart disease, limb damage (leading to amputation) and eye disease.

Diabetic retinopathy is a common complication of diabetes and a leading cause of blindness and visual impairment. After 15 years of diabetes, over 95% type 1 and 77% of type 2 patient sufferers develop retinopathy [5][6]. The severity of the retinopathy varies with the age of onset, disease duration and the blood glucose control. However, at 15 years duration, approximately 2% of people become blind and 10% will be severely visually handicapped [4]; see figure 1.1.

Although, diabetic retinopathy cannot be cured, the progression of the disease can be slowed or even halted if it is detected early and treatment given. Diabetics are generally unaware of the onset of retinopathy as visual loss is most commonly associated with severe retinal disease. It is therefore recommended that diabetics receive annual eye examinations to detect initial signs of progression of retinopathy. Such screening, followed by appropriate treatment is believed to prevent blindness

Figure 1.1: Effect on vision a). Normal vision b). Vision with diabetic retinopathy, [taken from http://www.stlukeseye.com/Conditions/DiabeticRetinopathy.asp]

within 90% of patients at risk if caught early enough in its progression [7].

Screening in the UK is either done by fundal examination (viewing the back of the eye, including the retina), performed by medical or optometric staff or by observation of retinal photographs [8]. Figure 1.2 a. shows a Canon CR6-45NM Non-Mydriatic fundus camera, the type used to capture retina images for this research. Figure 1.3 shows two examples of retinal fundal images. In photographic screening units highly trained staff visually assess large numbers of retinal images, searching for small abnormalities indicative of diabetic retinopathy.

Figure 1.2: Canon CR6 45MN, [taken from http://www.edigonline.com/fundus.htm]

As the growing diabetic population overstretches the screening infrastructure [8] the pressure on specialist staff increase. This can have a knock on effect for the detection of diabetic retinopathy as early indicators can be missed due to subjective error or human error in the performance of repetitive tasks. A screening method that does not require highly trained personnel would relieve pressure from screening services and decrease costs.

In summary, a fully automated diagnostic approach, involving computer analysis of digital fundus images, could provide patient referral assessments and reduce the workload of specialist staff.

1.2 Objectives

Automating retinal screening from digital fundus images can be thought of as a modular problem with varying degrees of success attainable. There are a number of visual clinical features indicative of diabetic retinopathy that appear at different phases of progression and at different levels of regularity. For an automated screening system to match a human grader, such a system would require an ensemble of algorithms capable of detecting and distinguishing all clinical features no matter how rare, or at least be capable of flagging unusual images.

The system would also be required to follow national screening guidelines, categorising patients as referable or non-referable. A patient at risk of developing sight threatening retinopathy and requiring further ophthalmic investigation is classed as referable whereas a patient not at risk and requiring no further action until the next annual eye examination is classed as non-referable. This classification depends upon the frequency and position of early diabetic features such as vascular bulges (microaneurysms), leakages of blood (haemorrhages), deposits of lipoproteins (exudates) and vascular anomalies such as venous beading (unusual variations in diameter).

In addition to accurately identifying diabetic lesions, locating structural elements such as the fovea, is also necessary. The fovea at the central part of the retina subserves fine detail and colour vision, retinal lesions found in, or immediately around, the fovea (macula) are associated with a high risk of visual loss as diabetic lesions or floaters obscure fine and colour vision. Figure 1.3a shows the structural location of the fovea together with the optic nerve head and vascular network. Figure 1.3b shows a high risk retina with diabetic retinopathy. The image has gross exudatation in conjunction with large haemorrhage floaters.

Identifying retinal lesions associated with diabetic retinopathy is a critical and demanding process in automated retinal image analysis. This is mainly due to the varying shapes, sizes, and colour of lesion types, image distracters such as lighting variation, natural pigmentation, light artefacts, broken capillaries and non-diabetic lesions. The variability of retinal appearance argues

for a structured approach [9], where significant landmarks in the retina are identified, and the relationships between them exploited to increase confidence in the classification of each object.

The focus of this research is the development of novel algorithms for the structured analysis of the retina, the purpose of which is the automatic diagnosis of digital images, identifying early signs of sight threatening retinopathy. As algorithms exist that segment the fovea and the blood vessel map to an acceptable level, this research has concentrated on the following areas of the structured analysis:

1. **Optic Nerve Head:** Optic nerve head segmentation is a necessary step in this structured analysis for a number of reasons. The optic nerve head can itself act as a distracter: it is a large bright region that can be mistaken (by algorithms) for gross circinate exudation (large circular group of exudates). Secondly, the optic nerve head can be used to locate the general location of the fovea using an approximated distance from the centre of the optic nerve. The fovea is a dark approximately circular area, but the contrast is often quite low, and it may be obscured by exudates or blurring. Consequently a global correlational search often fails.

2. **Diabetic Lesions:** Microaneurysms and haemorrhages detection and classification: Retinal microaneurysms are the earliest clinically apparent lesion indicative of diabetic retinopathy. This is shortly followed by the appearance of small haemorrhages that have leaked from ruptured microaneurysms. Identification of these lesions is therefore vital to obtain early indication of disease. Exudate detection and classification: Exudates located within the macula area are an important gauge of sight threatening retinopathy or progressive maculopathy (within the macula area).

3. **Image Quality Assessment:** Image quality is an important factor in diagnostic imaging, and can alter diagnosis if areas indicative of disease are blurred or not illuminated. Retinal image quality can be defined by the contrast between the fovea and background retina and the quantity and sharpness of retinal vessels leading up to the macula. Measuring image quality is a crucial step in order to flag images not suitable for automated screening.

4. **Retinal Vessels:** Changes in retinal vessel diameter are an important predictive indicator of sight threatening retinopathy. Obtaining precise measurements of vascular widths is therefore a necessary step to calculate vascular change.

The proposed automated system should be able to detect features symptomatic of diabetic retinopathy and provide diagnosis based on national screening criteria. It is expected that the majority of sight threatening retinopathy cases be identified and ungradeable images flagged for further investigation. The outline of the proposed system structure is shown in figure 1.4. The modular structure of the system allows continuous improvement to existing retinal analysis components, together with enabling the inclusion of additional clinical feature location and classification; such as the feature inclusion of venous beading detection.

1.3 Criteria for Success

As this research aims to investigate and develop techniques for the structured analysis and classification of digitized retinal images, its success can be judged on how well each algorithm fulfils its role and how it compares to other published techniques:

1. Locate and segment the Optic Nerve Head from the fundus image.
2. Segment the boundary of potential retinal lesions indicative of diabetic retinopathy.
3. Classify potential lesions into retinopathy and non-retinopathy categories.
4. Determine the image quality and assess gradability of image.
5. Vascular width measurements
6. Classify image into referable and non-referable maculopathy.

These criteria are revisited in chapter 8 and are compared against actual results.

1.4 Thesis Overview

This thesis is structured into a number of chapters, each focusing on a different aspect of background and research. The chapters are summarised as follows:

Chapter 2 - Medical Background: begins by describing the anatomy of the eye, followed by the fundamental causes of diabetes and the associated complications focusing on diabetic retinopathy. The pathological and clinical aspects of the eye disease are mentioned together with an overview

of the retinal treatment available. The chapter concludes with discussion on why early screening for diabetic retinopathy is important and why an automated screening system is preferable.

The structure of the following chapters include the Problem Domain and Previous Work, Method, Evaluation and Conclusion for their respective topics: Optic Nerve Head Segmentation, Lesion Detection and Classification, Image Quality Assessment and Vascular Diameter Measurement.

Chapter 3 - Optic Nerve Head: describes a novel algorithm for the localization and segmentation of the optic nerve head boundary in low-resolution images (about 20μ per pixel). Optic disk localization is achieved using specialized template matching, and segmentation by a deformable contour model. The latter uses a global elliptical model and a local deformable model with variable edge-strength dependent stiffness.

Chapter 4 - Fovea Detection & Vascular Segmentation: describes two algorithms presented by previous authors for fovea detection and vessel detection. The algorithms have been incorporated into the overall system and are described for completeness.

Chapter 5 - Lesion Detection: introduces techniques to automatically detect lesions suggestive of diabetic retinopathy. This is achieved using peak detection, novel region growing, feature selection and neural network classification.

Chapter 6 - Image Quality Assessment: describes a novel algorithm to quantify the gradability of retinal images. This is achieved by combining the fovea contrast with a weighted contrast between vascular centre pixels and the background.

Chapter 7 - Retinal Vessels: describes a new algorithm to measure the vessel diameter to sub-pixel accuracy. The diameter measurement is based on a two-dimensional difference of Gaussian model, which is optimized to fit a two-dimensional intensity vessel segment.

Chapter 8 - System Evaluation: demonstrates the performance of the integrated system (incorporating optic disc, fovea, vascular, and lesion segmentation together with image quality assessment) in detecting sight threatening diabetic retinopathy. The system evaluated on a per image basis against a variety of criteria and previously published work.

Chapter 9 - Conclusion: concludes the thesis by reviewing the presented research and discusses its novel contributions. The aforementioned criteria for success are also examined and are compared against actual results. Finally, areas for future work are described and their impact in the structured retinal analysis is highlighted.

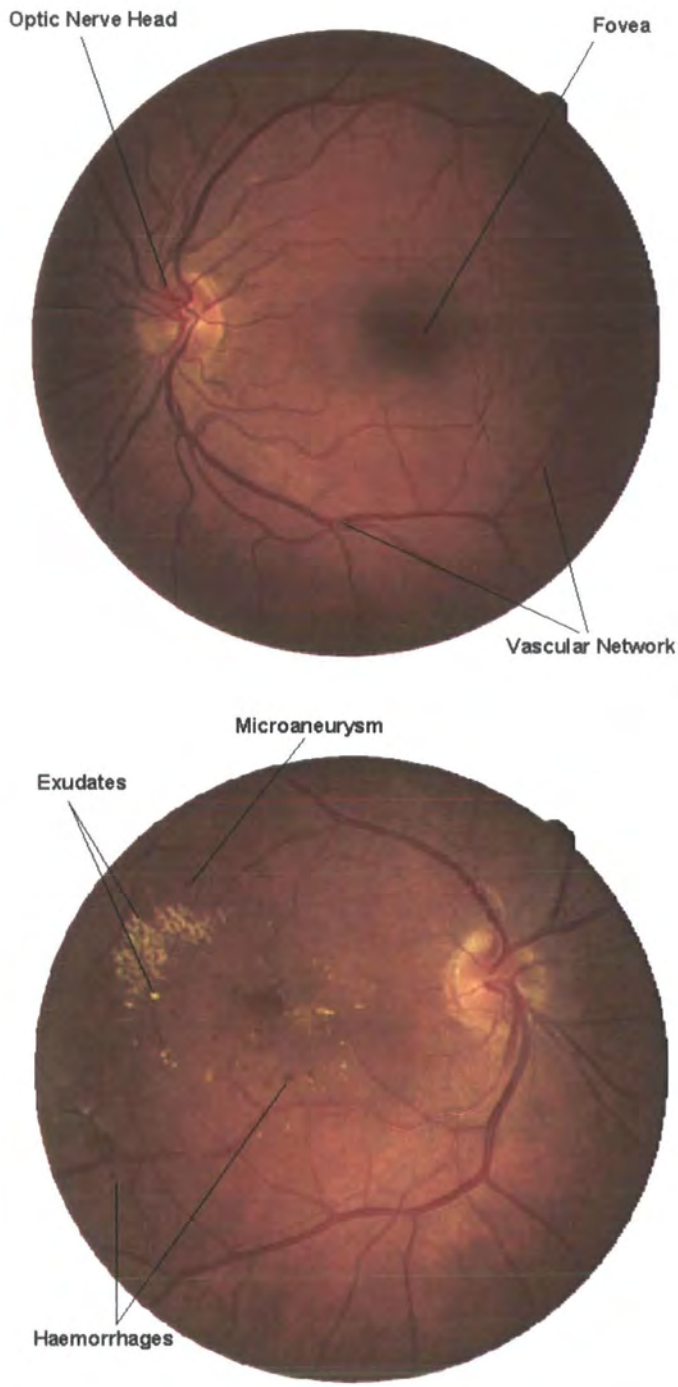


Figure 1.3: Fundus Photographs a). Structural components of the retina b). Retina with diabetic retinopathy

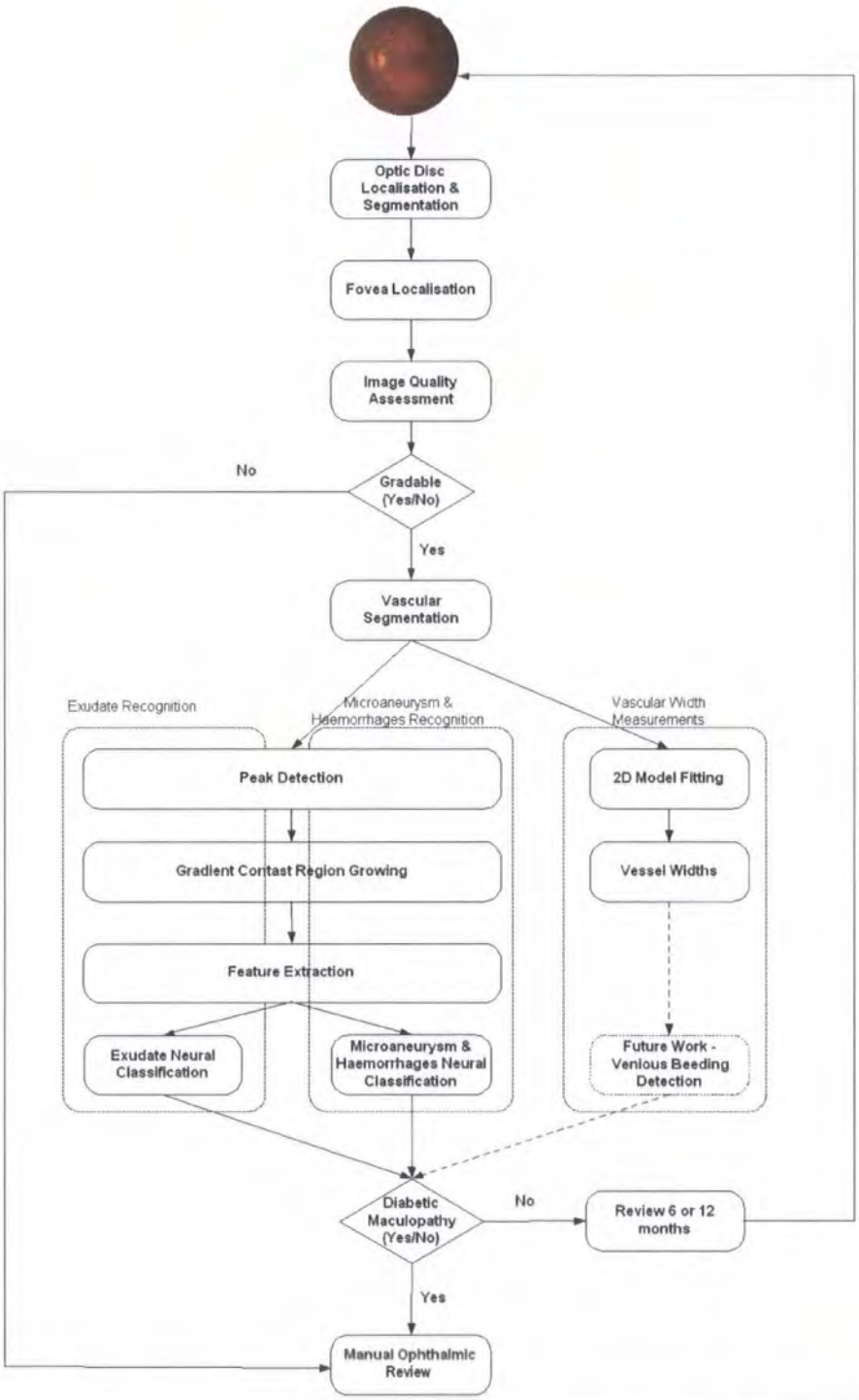


Figure 1.4: Outline of the proposed system for automated detection of maculopathy

Chapter 2

Medical Background

2.1 Introduction

Diabetes is a chronic disease affecting over 2% of the population in the UK [1]. Diabetes can manifest in different systems in the body causing long-term complications to the heart, kidneys, nerves, blood vessels and the eye. In the eye, diabetes can affect various ocular structures such as the lens and the retina. In the retina, diabetes can lead to an eye disease called diabetic retinopathy, which has been identified as one of the leading causes of blindness in the working population of industrialised countries. Persons with diabetes are 29 times more likely to become blind than those without the disease [10].

As the initial phase of the disease is typically asymptomatic, individuals are not usually aware of the risk of developing sight-threatening retinopathy and consequently loss of vision until it is too late. It is therefore crucial that regular retinal examinations are carried out in order to detect the initial onset of the disease before noticeable visual loss occurs.

To reduce the risk of visual loss from diabetic retinopathy, treatment is required at the onset of sight-threatening retinopathy. Early treatment by laser photocoagulation has been shown to significantly reduce the incidence of visual loss [11].

Diabetic retinopathy can be detected by either clinical examination using different methods such as direct ophthalmoscopy, indirect ophthalmoscopy and biomicroscopy; or retinal photography using instant fundus Polaroid photographs, colour fundus 35-mm slides or digital imaging. The

UK National Screening Committee have suggested that annual screening by digital colour fundal photography is the preferred medium [2]. With 20,000 - 30,000 diabetic patients per million population, manually grading retinal images is a time-consuming and costly process requiring highly skilled staff and is susceptible to subjective variation and error.

This chapter begins with a brief description of the anatomy of the eye, followed by a discussion of the causes and pathological changes of diabetic retinopathy. The clinical features suggestive of diabetic retinopathy are described followed by current methods of diagnosis and treatment of the disease. The chapter is concluded by emphasizing the benefits of automated retinal analysis for digitised retinal photography.

2.2 Anatomy of the Eye

The eye is a complex optical structure that is able to reflect and focus light that stimulates neural responses enabling us to see. The eye is essentially made up from a number of optical components, neural components and supportive layers, see figure 2.1. At the front of the eye is a thin transparent membrane known as the cornea. The cornea has a dual purpose of protecting the eye, and refracting light as it enters the eye. A portion of the light passing through the cornea passes through the pupil, a small opening in front of the lens.

The amount of light that can pass through the pupil is dependant on the size. This is adjusted by contractions of the iris muscles. The iris is a diaphragm that is capable of enlarging and reducing the size of the pupil. In bright light, the iris constricts, reducing the size of the pupil and the amount of light entering the eye. In dim-light the iris dilates, enlarging the pupil size and increasing the amount of light entering the eye. The light then passes through a clear crystalline lens. The lens controls 1/3 of the refraction of light that enters the eye (the cornea, the other 2/3) completing refraction by fine-tuning the focused light onto the retina.

The eye is full of a clear gelatinous substance known as vitreous gel that constitutes 2/3 of the eye volume and it is formed mainly of water and structural proteins. Light refracted from the lens must travel through the vitreous before reaching the retina.

The retina is the innermost membrane that lines the back of the eye. It comprises of multi-layered sensory and pigmentary layers. The retina contains photoreceptions that transform cap-

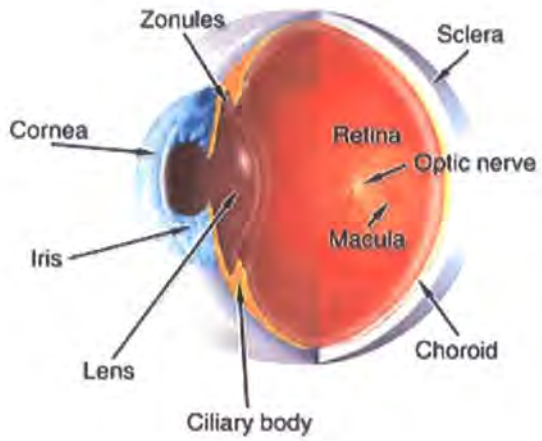


Figure 2.1: Anatomy of the eye, [taken from http://www.stlukeseye.com/Anatomy.asp]

tured light rays into electrical nerve impulses. These impulses travel from the light photoreceptors of the retina along the optic nerve and into the brain's visual cortex for processing. There are two types of photoreceptors in the retina: rods and cones. The cones, which number approximately 6 million, are primarily located in the central part of the retina called the macula. Cones are highly sensitive to colour, functioning best in bright light and enable colour vision.

The highest density of cones is situated at the centre of the macular called the fovea. The fovea is responsible for central, sharp vision and this is largely due to each cone being connected to its own nerve fibre. The fovea itself is a circular indentation in the retina of approximately 1.5mm in diameter [12] and unlike the peripheral retina the fovea has no blood vessels. See figure 2.2

Approximately 75 to 150 million rods are distributed throughout the peripheral retina. As several rods are connected to single nerve endings the information discernable from these receptors is reduced giving peripheral vision. Rods are sensitive to low levels of illumination and receive no colour information; as a result night vision is colourless.

2.3 Causes of Diabetic Retinopathy

Diabetes mellitus is a chronic disease triggered by a metabolism disorder resulting in an unhealthy level of glucose in the blood. The digestion of starchy foods such as rice, potatoes, bread and sugar

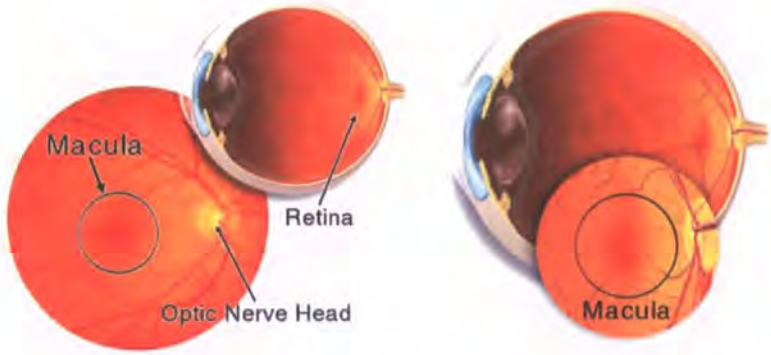


Figure 2.2: Anatomy a). Retina. b). Macula, [taken from http://www.stlukeseye.com/anatomy /retina.asp and /Macula.asp respectively]

produces glucose that enters the bloodstream where it is used by cells for growth and energy. For glucose to get into the cells the hormone insulin must be present. When the pancreas does not secrete enough insulin, the cells do not respond appropriately and glucose builds up in the blood. The body therefore loses its main source of fuel as it cannot use the glucose present in the blood.

The long-term complications of diabetes affect almost every part of the body. The disease often leads to heart disease, kidney failure, nerve and blood vessel damage. Damage to the small blood vessels throughout the body can lead to reduced blood flow. When these changes affect the tiny blood vessels in the retina of the eye, diabetic retinopathy may develop.

In the early stages of diabetic retinopathy, the tiny retinal vessels weaken and develop outpouching (microaneurysms), which may burst and leak blood (haemorrhages) and fat (exudates) into the retina. The leakage may also cause swelling (oedema) of the retina. The weak blood vessels may also become blocked, starving the retina of nourishment, leading to the growth of new abnormal vessels in the retina. The new blood vessels are often fragile and can bleed into the eye or may form scar tissue that pulls on the retina, leading to retinal detachment.

2.4 Pathological Changes of Diabetic Retinopathy

Diabetic retinopathy can be pathologically classified into three principal categories: non-proliferative, proliferative retinopathy and maculopathy. Non-proliferative retinopathy can be further classified

into three different stages according to its severity.

2.4.1 Non-proliferative Retinopathy

1. **Mild Non-proliferative Retinopathy.** At the earliest stage, balloon-like swelling occurs in the retina's blood vessels. These swellings are called microaneurysms and may not have any noticeable effect on vision.

2. **Moderate Non-proliferative Retinopathy.** Disease progression results in a number of blocked blood vessels reducing the nourishment to the retina. The weakened capillaries can lead to retinal haemorrhages, fluid and leakage lipid accumulation (exudates).

3. **Severe Non-proliferative Retinopathy.** An increased number of blood vessels become blocked, depriving areas of the retina with their blood supply. The deprived retina secretes chemicals such as growth factors; these chemicals can trigger the growth of abnormal new vessels.

2.4.2 Proliferative Retinopathy

At this advanced form of the disease, new, abnormally weak, blood vessels grow on the retinal surface or overlying the optic disk; these vessels extend in the plane between the retinal surface and the posterior surface of the vitreous gel (the clear jelly-like substance that fills the centre of the eye) and acts as a scaffold for its further growth. The abnormal growth of new blood vessels does not support deprived areas of the retina with a new blood supply, instead these vessels may cause other complications:

1. **Vitreous haemorrhage.** Subsequent contraction of the vitreous gel leads to traction on the fragile new blood vessels that can lead to their avulsion. This results in bleeding either in or behind the vitreous gel with resultant drop of vision. If the amount of bleeding is small only a few dark spots may obscure vision. In severe cause the entire vitreous cavity can fill with blood and subsequently block all vision.

2. **Traction retinal detachment.** Scar tissue accompanies new blood vessel growth. This scar tissue can contract and pull the retina from the back wall of the eye. This can result in profound visual loss if the macula is involved.

2.4.3 Maculopathy

Maculopathy can occur in both non-proliferative and proliferative retinopathy and is the commonest cause of visual loss within the diabetic population. Diabetic Maculopathy can be defined as diabetic retinopathy affecting the central macula [13]. The condition can be divided into several subgroups each with a different pattern of disease; focal, diffuse, ischaemic and mixed.

1. **Focal** maculopathy consists of clusters of microaneurysms from deep or superficial capillary networks associated with focal retinal oedema, and often surrounded by hard exudates. This is due to lipoproteins (seen as hard exudates) being deposited at distal sites where leakage causes a flow of fluids towards normal capillaries for resorption.

2. **Diffuse** macular oedema occurs from widespread blood-retinal barrier break down and diffuse dilatation of the capillary bed. Hard exudates may not be present and microaneurysms often inconspicuous.

3. **Macular** ischaemia occurs from capillary closure, often causing deep retinal haemorrhages and cotton wool spots representing nerve fibre layer infarcts and occasionally visibly closed white vessels.

4. **Mixed** maculopathy contains a combination of all three groups with the full extent visible only on fluorescein angiography [13]. Only focal, diffuse and mixed maculopathies can be treated with laser photocoagulation.

2.5 Clinical Features

As the retinopathy progresses a small number of clinical features appear in the retina. Their characteristics and position in relation to the macula and each other is of great importance when making a diagnosis. There are two types of clinical feature suggestive of diabetic retinopathy: lesions and vascular anomalies. Retinal lesions include haemorrhages and microaneurysms (dark spots), cotton wool spots and exudates (bright spots), and vascular features including venous beading.

1. **Microaneurysms.** Retinal microaneurysms are the earliest clinically apparent lesion indicative of diabetic retinopathy. The number of microaneurysms increases with the severity

of the retinopathy [14]. A microaneurysm indicates an area of deficient capillary cell wall and localised break of the blood-retinal barrier and may be associated with abnormal vascular leakage [15].

2. **Haemorrhages.** It is thought that haemorrhages occur from ruptured microaneurysms or weak vascular segments. Haemorrhages vary in appearance more than microaneurysms and are not necessarily round. Haemorrhages can occur in the confinement of the retina or they can spread out on the retinal surface where they take on a flame appearance [15].

3. **Exudates.** Exudates are frequently associated with clinically significant macular oedema. Previously called hard exudates to distinguish them from soft exudates (now called cotton wool spots), exudates are collections of lipoproteins caused by abnormal vascular leakage. The number of exudates increases as extravascular fluid diminishes due to precipitation of lipids and proteins; this is analogous to a saline solution depositing salt upon drying [15].

4. **Cotton Wool Spots.** Cotton wool spots are areas of swollen nerve axons caused by a localised obstruction of blood supply in the nerve fibre layer. [15]. They appear as small whitish fluffy superficial lesions that obscure underlying blood vessels and are clinically evident only in the post equatorial retina where the nerve fiber layer is of sufficient thickness to render them visible [16].

5. **Venous Beading.** Venous beading is an important sign of sluggish retinal circulation. It occurs in the retinal veins in a response to oxygen deficiency, it represents a localised increase in the venous calibre (segmental dilatation), in severe cases it appears as a string of beads 'sausage-like appearance'. Venous beading is indicative of severe non-proliferative retinopathy and is the most significant predictor of proliferative diabetic retinopathy, ie: imminent new vessels development. [15] [?].

Non-diabetic lesions that are commonly mistaken as Exudates due to their similar visual characteristics include Drusen. Drusen are nodules that exist in a layer called Bruch's membrane which lies beneath the retina and the adjacent retina pigment epithelium layer. Drusen are metabolic waste derived from retina pigment epithelium cells RPE. Its accumulation is thought to result from failure to clear the debris discharged into the region. Drusens appear as yellow excrescences

beneath the RPE distributed symmetrically at both posterior poles. They can vary in number and size. Drusens are rarely clinically visible before the age of 45 years [17] [16].

2.6 Clinical Diagnosis

The diagnosis of diabetic retinopathy is based upon the presence of the various clinical features previously described. The clinical criteria as set by the Early Treatment Diabetic Retinopathy Study Group [11] are used to classify the extent of diabetic retinopathy:

2.6.1 Non-proliferative Diabetic Retinopathy (NPDR)

1. **Mild NPDR**: Microaneurysms with or without one or more of the following: Intra-retinal hemorrhages, hard exudates away from macula or cotton wool spots

2. **Moderate NPDR**: Microaneurysms/hemorrhages in at least one quadrant (see figure 2.3) plus one or more of the following: cotton wool spots, intra-retinal microvascular abnormality or venous beading.

3. **Severe NPDR**: Any one of the following (4-2-1 rule): Intra-retinal hemorrhages in 4 quadrants, venous beading in 2 quadrants and moderately severe intra-retinal microvascular abnormality in 1 quadrant.

2.6.2 Proliferative Diabetic Retinopathy (PDR)

1. **Early PDR**: One or more of the following: Neovascularisation of the optic nerve head (also known as optic disc) $< \frac{1}{4}$ disc diameter. Neovascularisation elsewhere (non-bleeding). Pre-retinal or vitreous hemorrhage and neovascularisation elsewhere $< \frac{1}{2}$ disc diameter without neovascularisation of the optic disc.

2. **High Risk Characteristics (HR PDR)**: One or more of the following: Neovascularisation of the optic disc $> \frac{1}{4}$ disc diameter. Neovascularisation of the optic disc with bleeding. Neovascularisation elsewhere $> \frac{1}{2}$ disc diameter with bleeding.

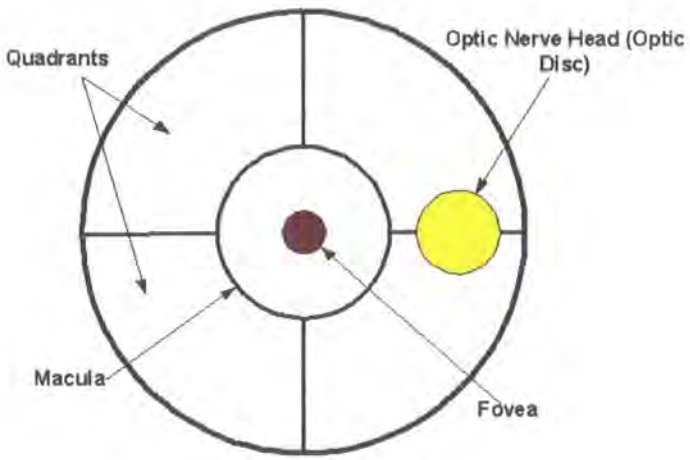


Figure 2.3: Retinal Quadrants

2.7 Treatment

Currently there are two forms of treatment for diabetic retinopathy: photocoagulation and vitrectomy. Both treatments can be effective in slowing or stopping the progression of the disease for a time. As diabetes continually weakens the arterioles and capillaries further retinal damage may occur and additional treatment is required. The decision of which procedure to use depends on the type and severity of the retinopathy.

2.7.1 Photocoagulation

In Photocoagulation, a high-energy laser beam creates small burns in the retina areas with abnormal blood vessels. The purpose of photocoagulation is to stop leakage of blood and fluid into the retina and hence slow the disease's progression. This can be achieved using one of two photocoagulation techniques: focal or panretinal.

1. **Focal Photocoagulation.** Diabetics with clinically significant macula edema, which is related to leaky capillaries and microaneurysms, are generally recommended to undergo focal photocoagulation laser treatment. In this procedure, the laser is applied to the macula of the eye avoiding the fovea and focused on leaking blood vessels creating "spot welds" to stop the leakage [18]. If the leaks are small in number the treatment is applied directly to specific leaks otherwise laser burns are applied in a grid-pattern to cover a wider area.

For diabetics diagnosed with clinically significant macula edema the risk of visual loss is reduced by more than 50% when treated by focal photocoagulation; see figure 2.4 a.

2. **Panretinal Photocoagulation.** For proliferative retinopathy laser panretinal photocoagulation is started promptly after diagnosis of the condition. In this treatment the entire retina except the macula is exposed to randomly placed laser burns, causing the new abnormal blood vessels to regress.

   Although the risk of vitreous haemorrhage and traction retinal is reduced, some loss in peripheral vision may occur. Panretinal photocoagulation is a trade-off, sacrificing peripheral vision to preserve central vision; see figure 2.4 b.

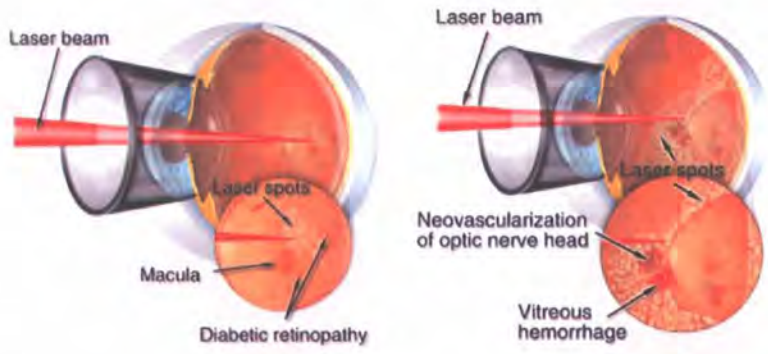


Figure 2.4: Photocoagulation a). Focal b). Panretinal, [taken from http://www.eyemdlink.com/ EyeProcedure.asp?EyeProcedureID=55]

2.7.2 Vitrectomy

If proliferative retinopathy is diagnosed late, one of the abnormal neovessels (neovascular vessel) may bleed into the vitreous, blocking vision and preventing laser treatment as the blood also blocks the laser. If the haemorrhage does not clear by itself within a few weeks or months, vitrectomy surgery may be performed. See figure 2.5.

In this procedure a surgeon removes the blood-filled vitreous from the eye, the abnormal new vessels are then dissected. A vitrectomy may also be indicated for the treatment of tractional retinal detachment. In this case the scar tissue that is pulling the retina away from the eye wall is dissected allowing the detached retina to flatten out and reattach itself.

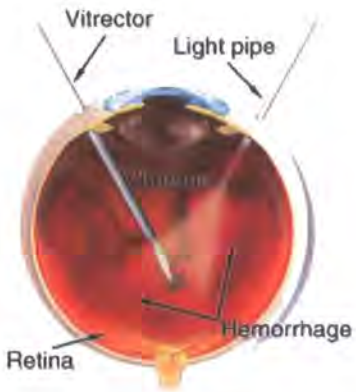


Figure 2.5: Vitrectomy, [taken from http://theretinasource.com/conditions/diabetic_ retinopathy.htm]

2.8 Early Detection for Diabetic Retinopathy

Treatment of diabetic retinopathy can prevent blindness within 90% of patients at risk, if the disease is caught early enough in its progression [7]. Annual screening is therefore necessary to detect those at risk before visual loss occurs and when treatment is at its most effective.

At present annual screening for diabetic retinopathy is carried out by general practitioners, optometrists, diabetologists, nurse practitioners, junior physicians and ophthalmologists. The level of experience and competence between and within these groups has been shown to vary significantly.

Automating the retinal analysis process to reduce the subjective variation and error would be beneficial to standardise screening. Such a system could have a number of roles in national screening:

1. **Screening tool**: highlighting potential lesions with or without classification or overall referable, non-referable classification.

2. **Audit tool**: Assessing the quality of national screening.

3. **Automated screening**: producing referable, non-referable recommendations, removing the need for national screeners. According to national guidelines any screening system should have sensitivity and specificity over 80% and 95% respectively, where sensitivity represents the percentage of true diabetic retinopathy cases detected and specificity represents the

percentage of non-diabetic retinopathy cases detected.

2.9 Summary

Long-term complications of diabetes can affect different systems of the body including the eye. In the eye, diabetes can affect the retina and lead to an eye disease called diabetic retinopathy which has been identified as one of the leading causes of blindness in the UK. As retinopathy progresses, tiny retinal vessels weaken and may rupture causing a leakage of blood or lipoprotein that may obscure central vision (macula) and may also cause swelling of the retina. The weakened blood vessels may also become malnourished leading to the growth of new fragile vessels that can bleed into the eye and again obscure vision.

Early detection followed by treatment can significantly reduce the risk of blindness. At present diabetic retinopathy is detected manually by either clinical examination or retinal photography. The UK national screening committee have suggested that annual screening by digital photography is the preferred medium. Automating retinal analysis using digitised retinal photography would be beneficial in reducing subjective variation and error and could have a number of roles in national screening; as a screening tool, audit tool or for fully automated screening.

Chapter 3

Optic Nerve Head

3.1 Introduction

This chapter presents a novel algorithm for the automatic localisation and segmentation of the optic nerve head in retinal images. The algorithm automatically selects the general location of the center of the optic nerve head, then fits a contour to the optic nerve head rim. Localization is achieved using a simple but effective specialized filter; segmentation by fitting an active contour to the optic nerve head rim using a three phase global and local deformable model that exploits the specific characteristics of the optic nerve head's appearance. The performance of the algorithm is evaluated against alternative approaches using a set of 100 random images drawn from a diabetic screening programme.

3.1.1 Motivation

Optic nerve head segmentation is a necessary step in this structured analysis for a number of reasons. First, the optic nerve head can itself act as a distractor: it is a large bright region that can be mistaken (by algorithms) for gross circinate exudation; the high-contrast rim also causes false responses to linear blood vessel filters [9]. Second, the vessels radiate from the optic nerve head, so vessel tracking algorithms [19] may start from there. Vessels are of direct importance in assessing vascular condition [20] [21] [22], (explained in more detail in chapter 7). Third, the optic nerve head is important in localisation of the fovea, the central part of the retina that subserves

Chapter 3

Optic Nerve Head

3.1 Introduction

This chapter presents a novel algorithm for the automatic localisation and segmentation of the optic nerve head in retinal images. The algorithm automatically selects the general location of the center of the optic nerve head, then fits a contour to the optic nerve head rim. Localization is achieved using a simple but effective specialized filter; segmentation by fitting an active contour to the optic nerve head rim using a three phase global and local deformable model that exploits the specific characteristics of the optic nerve head's appearance. The performance of the algorithm is evaluated against alternative approaches using a set of 100 random images drawn from a diabetic screening programme.

3.1.1 Motivation

Optic nerve head segmentation is a necessary step in this structured analysis for a number of reasons. First, the optic nerve head can itself act as a distractor: it is a large bright region that can be mistaken (by algorithms) for gross circinate exudation; the high-contrast rim also causes false responses to linear blood vessel filters [9]. Second, the vessels radiate from the optic nerve head, so vessel tracking algorithms [19] may start from there. Vessels are of direct importance in assessing vascular condition [20] [21] [22], (explained in more detail in chapter 7). Third, the optic nerve head is important in localisation of the fovea, the central part of the retina that subserves

fine vision.

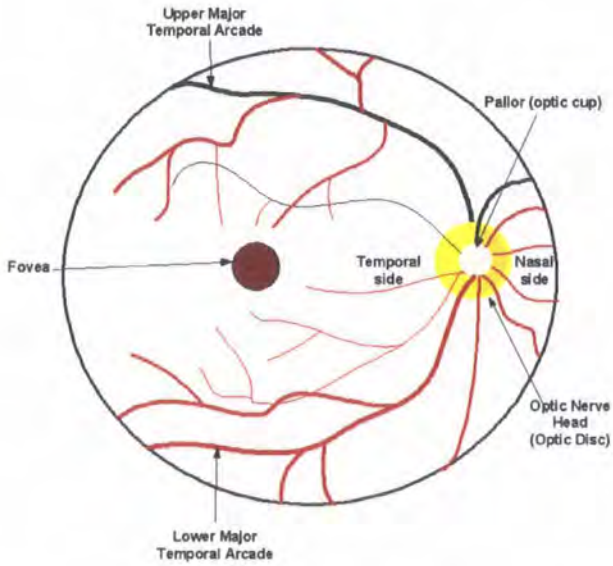


Figure 3.1: Retinal Structure

In fundus images where the fovea is obscured by lesions or has poor contrast, it is necessary to approximate its location in order to identify high-risk lesions near and around the fovea. The fovea is located approximately 2.5 optic disc diameters temporal to the temporal edge of the optic disc and lies between the major temporal retinal vascular arcades (see figure 3.1. Using the positional constraints, the approximate location of the fovea can be calculated and used as a substitute value if the foveal search methods fails, although variation in the optic disc size compromises the reliability of this method. Littmann [23] has developed a technique to explicitly determine the distance from the optic disc's center to the fovea, by correcting for the magnification factors of the fundus camera and the patient's eye. However, the ocular magnification factor depends mainly on anterior corneal curvature, refraction and axial length, which require measuring variables on the patient's eye – an important practical limitation in screening. The method can also give false values in the case of an abnormally high lens refractive power (e.g. by cataract formation) and so is not applicable in aphakic or pseudophacic eyes (10% of the diabetic screening population). The "2.5 disc diameters" heuristic method is therefore more practical as absolute accuracy is not required.

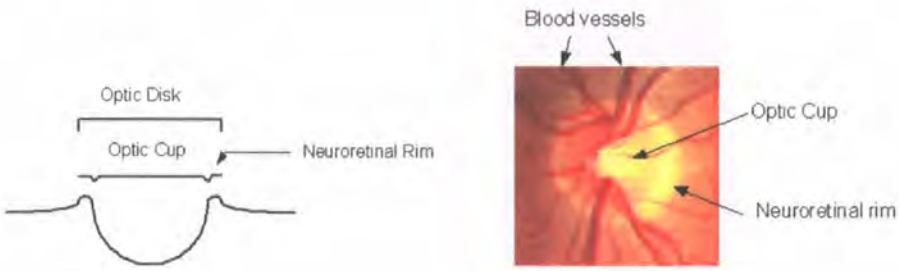


Figure 3.2: The Optic Nerve Head. a) Cross Section. b) A typical well-defined disc.

3.1.2 Chapter Contents

Section 3.2 describes the appearance and complexities of optic nerve head localisation and segmentation. Two types of boundary segmentation are described - ellipse fitting using hough transforms and active contour modelling (also known as snakes), together with previously published optic nerve head location and segmentation algorithms. In section 3.3, the optic nerve head location and boundary segmentation algorithms are described. The segmentation algorithm is evaluated against alternatives from the literature in section 3.4 and concluded in section 3.5.

3.2 Problem Domain & Literature Review

3.2.1 Optic Nerve Head Appearance

Successful segmentation of the optic nerve head requires a careful analysis of its appearance (see figure 3.2). It is the extremity of the optic nerve in the interior of the eye, and also the entrance and exit site of retinal arteries and veins [24]. The shape of the optic nerve head is approximately elliptical, with a vertical principal axis (width $1.8 \pm 0.2mm$, height $1.9 \pm 0.2mm$) [25]. As the nerve fibres reach the optic nerve head they turn and exit through the optic nerve, leaving a small depression (the "cup") in the center of the nerve head. There is often a brighter central region, the "pallor", which if present usually includes the cup. The optic disc rim is judged to be the inner margin of the peripapillary scleral ring, seen as a thin white band encircling the optic disc.

In fundal images, the appearance varies quite substantially; see figure 3.3. Although the topology of the optic disc is standard, there are large variations in the size and shape of the optic nerve

head, due to variable amounts of glial and connective tissue. The rim is usually visible as a bright boundary with the nasal side usually being less bright than the temporal side, however sometimes the rim is not visible at all. In some images the entire optic nerve head is brighter than the surrounding area, appearing as a disc; in others the appearance is of a hollow ring. In either case the pallor may appear as a smaller, brighter disc within the optic disc. There may also be bright areas just outside the rim caused by peripapillary atrophy, either distorting the shape or forming concentric elliptical arcs. To complicate the issue further, departing vessels partially obscure the rim. The majority exit on the nasal side and depart vertically; a smaller number depart nasally, and a few fine vessels depart temporally. Occasionally vessels turn at the nasal rim edge and run vertically, obscuring portions of the optic boundary. A consequence of the nasal distribution of vessels is that the pallor, if visible, is mainly located to the temporal side.

The variability in appearance misleads obvious localisation and segmentation approaches. Large areas of bright diabetic lesions (gross exudation see section 5.2.1), which have high contrast, act as strong distractors for correlation-based localization algorithms – algorithms that work well on images of healthy retina may fail on a screening population. Similar problems arise from reflection artifacts (young patients) and visible choroidal vessels [26].

Segmentation is complicated by the presence of strong distractors along the pallor and vessel edges, weakness of the rim and peripapillary atrophy [27]. However, these problems can be overcome by exploiting specific aspects of the appearance: the relative sharpness, reliability and lack of vascular intersections on the temporal side of the rim, and the approximately elliptical shape.

3.2.2 Optic Nerve Head Localisation

Optic Disc Localisation Algorithms

Reliable optic nerve head location is surprisingly difficult, due to its variable appearance; naïve approaches that work well for images of healthy retinae often fail on screening images.

Sinthanayothin *et al.* [25] exploited the rapid intensity variation between the dark vessels and the bright nerve fibres to locate the optic disc. This was achieved by using a 80 x 80 sub-image to calculate the intensity variance of adjacent pixels. An average variance within these sub-images was obtained and the point of largest average variance deemed the optic centre. The sensitivity and specificity of this technique is reported as 99.1% when no significant distractors are present. When

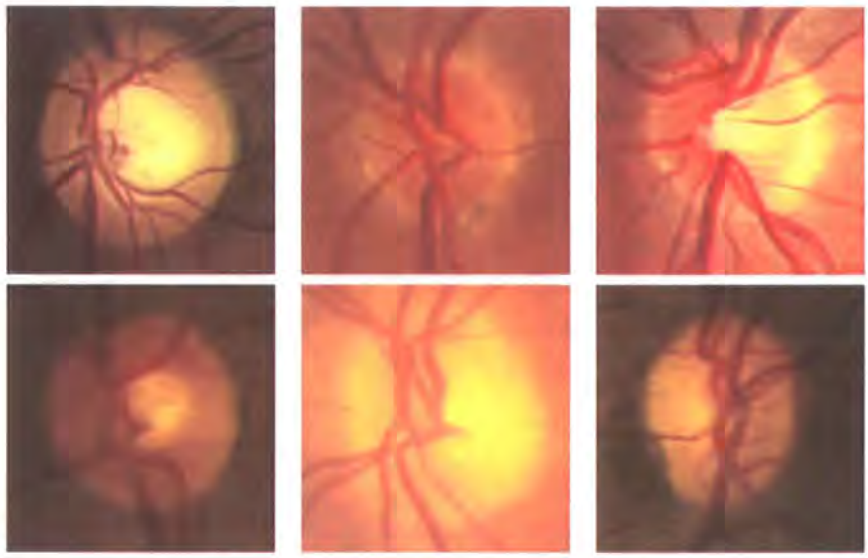


Figure 3.3: Examples of varing optic nerve head appearance

presented with large cotton wool spots, gross exudation, light artifacts or strongly visible choroidal vessels the algorithm commonly misclassifies the optic disc's location for that of a distractor.

Akita *et al.* [24], traced the parent-child relationship between blood vessels segments, tracking back to the center of the optic disc; this presupposes robust detection of the blood vessels, which is difficult in images of diseased retinae where even quite sophisticated algorithms detect false positives along the edges of white lesions and along the edge of the optic nerve head itself [9].

Lalonde *et al.*[28], used pyramidal decomposition and Hausdorff-based template matching to locate the optic disc. The template-based matching technique is based on a edge map using a Hausdorff distance measure and guided by scale tracking of large objects using multiresolution image decomposition. This method is effective, but computationally demanding.

3.2.3 Optic Nerve Head Boundary Segmentation

Overview of Circular Hough Transform

The Hough transform is a global, robust technique for the detection of predefined shapes in images [29]. Since patented in 1962, the technique has been successfully employed in a range of domains including the detection of human hemoglobin fingerprints [30], the detection of tumors in chest

films [31] and more recently optic nerve head segmentation.

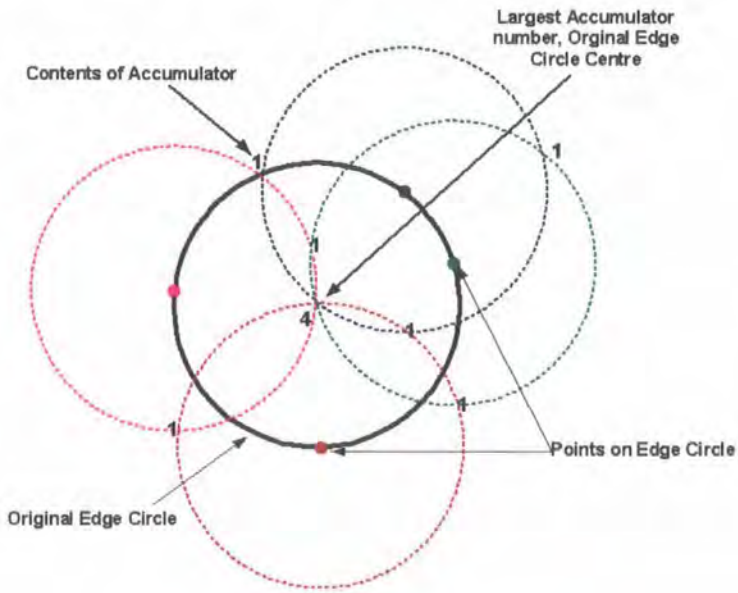


Figure 3.4: Circular Hough Transform

The Hough transform works by converting edge points to parameter space and exploiting the symmetry of parametric shapes such as lines, circles and ellipses. The transform assumes that some form of pre-processing has taken place to create edge features. A range of algorithms such as Canny edge detection or zero crossing of the laplacian can extract this edge description. Obtaining good edge descriptions is vital for the success of the Hough transform; domain issues usually dictate the appropriate edge detection algorithm.

Although originally designed for the detection of straight lines, any parametric shape can be used. In theory, this method can be used to find features of any shape in an image. However in practice it is only commonly used for simplistic shapes such as straight lines, circles, and ellipses, as more complex shape dramatically increase the computational complexity.

In circular Hough transforms, each pixel from the edge feature map centres a circle with a radius of r. Figure 3.5b. shows an example of an optic edge feature map. The point at which most circles intersect is deemed the center of the feature edge circle. The circular Hough transform works by considering all circles centred at each edge feature pixel at once and identifying co-ordinates in Hough space where circles intersect. The frequency of these intersects at each co-ordinate are

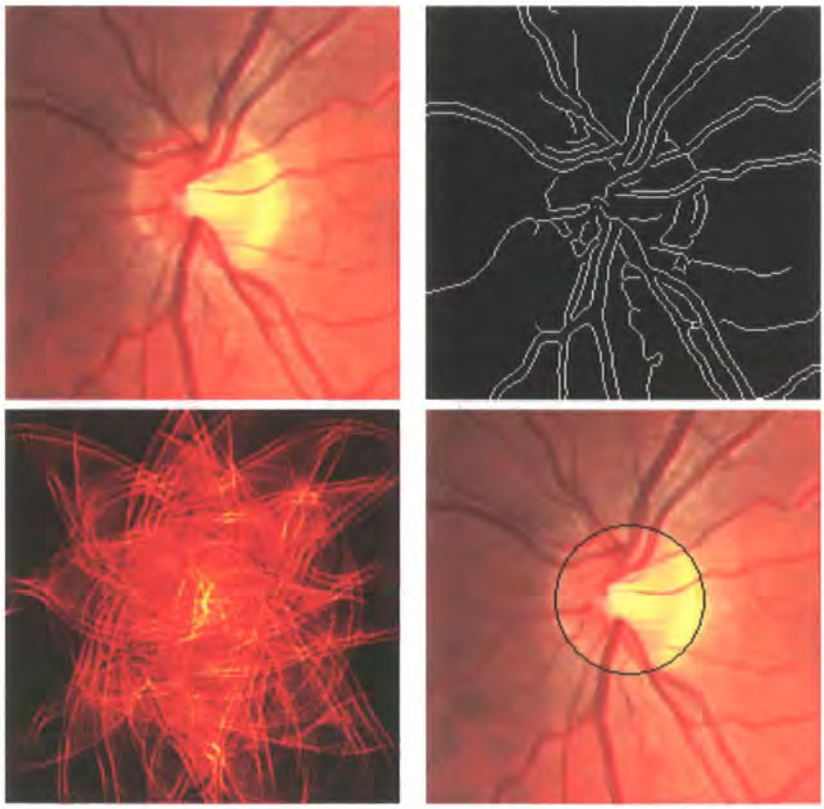


Figure 3.5: Circular Hough Transform a)Colour fundus. b)Canny edge map. c)Circular Hough Transform. d)Superimposed Hough Result on Optic Nerve Head.

totalled in an accumulator space (see Figure 3.4). The dimension of the accumulator is determined by size of the parametric representation of the shape.

The co-ordinate in the accumulator with the highest frequency of intersecting circles is defined as the centre of the edge feature circle (see figure 3.5c). If the size of the circle is unknown an additional parameter is required - radius, thus the circle can be parameterised by (x,y,r) and the accumulator space is increased to three dimensions. An example of a superimposed Hough result is shown in figure 3.5d.

Optic Nerve Head Boundary Segmentation using Hough Transform

Lee [32] experimented with elliptical Hough transforms to extract the optic boundary from optic centred images. These images were converted to grayscale prior to Canny edge detection. The

resultant edge maps although roughly circular were broken due to crossing blood vessels and varied contrast. To increase the efficiency of the Hough transform, Lee reduced the size of the parameter space using prior knowledge of the size and position of the optic nerve head. Lee reported mixed results from the elliptical Hough transform.

Ege et al. [33] used Sobel edge detection to define optic edge features and calculated the image gradient to utilise directional information. The circular Hough transform was applied to gradient images to locate the optic boundary.

The main advantage of the Hough technique is that it is relatively unaffected by noise or gaps in the edge feature. There is however a point at which poor edge detection will yield poor results or even fail. With this in mind it should be noted that low contrast, and blurred boundaries commonly cause optic disc edge detection to provide unacceptable solutions; hitting either curved blood vessel segments or the pallor boundary. In addition, it is not uncommon for the optic disc to have irregular boundaries that cannot be accurately modelled by the Hough transform. This would result in an un-segmented rim around the optic nerve and could cause false exudate detection (see Figure 3.6).

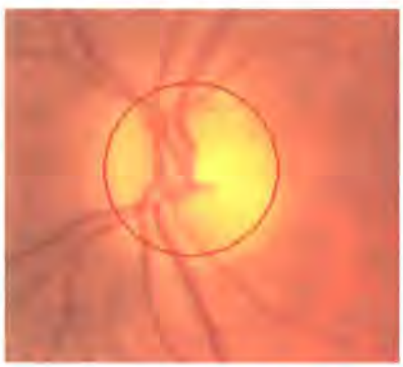


Figure 3.6: Hough Transform Result

Overview of Deformable Contour

Active contours, or snakes [34], are deformable models that are fit to object edges under the control of two forces: *external forces* that pull the model towards image features such as edges; and *internal forces* that act as smoothing constraints or object model constraints. There are two main categories: freeform and parametric snakes. Freeform snakes provide many local degrees of freedom, whereas

parametric deformable models encode a specific shape and can help to overcome image problems such as boundary gaps [35] [36]. The optic nerve head is well-suited to the parametric approach, as it has a simple global model and significant distractors.

Active contours seek points with a high gradient. There are two main problems: noise, which may trap the model in local minima; and the domain of edge attraction – if too small, the snake may not find the desired solution [37]. These issues are addressed by pre-filtering with a low-pass Gaussian filter [38]. Often the gradient magnitude image, $\|\nabla I\|$, is used; in gradient vector methods the direction of the vector is retained to help locate edges with an expected orientation.

Optic Nerve Head Contour Algorithms

There are several key problems in using active contours for optic nerve head segmentation. Firstly, blood vessels contribute powerful distractors along their edges, and obscure parts of the optic boundary. The pallor edge (optic cup) may present a strong contrast boundary; it may also combine with the temporal edge of vessels on the nasal side to form a strong elliptical distractor. There may also be bright concentric arcs of peripapillary atrophy outside the rim. The rim particularly on the nasal side may be blurred or hidden by blood vessels. Two previous authors have reported the use of active contours to find the optic disc boundary.

Mendel [39] used a freeform snake, initialized as a circle centered on and inside the optic disc. Initially, the model tended to fit a convoluted boundary, following vessel edge distractors. Mendel addressed this problem using grey-scale morphological closure to remove blood vessels from within the optic disc. Figure 3.7 shows an optic disc before and after morphological closure. Mendel evaluated nine high quality images. Due to the lack of a global model it is expected that the algorithm would perform poorly on blurred images or images with a low intensity rim or significant distractors at the pallor edge. Mendel used a gradient vector method, and noted that this helps to avoid distractors.

Lee [32] also applied an active contour model to high resolution images centered on the optic nerve head. Like Mendel, he removed vessels morphologically. The control points and attractors were placed along radial spokes emanating from the center, thus partially imposing a global model. Attractor points were attached to edges detected using a Canny filter, which makes the system prone to complete failure where the rim is missing. Lee reported problems caused by the boundary

of the pallor and by very faint or missing edges, and presented results on only four images.

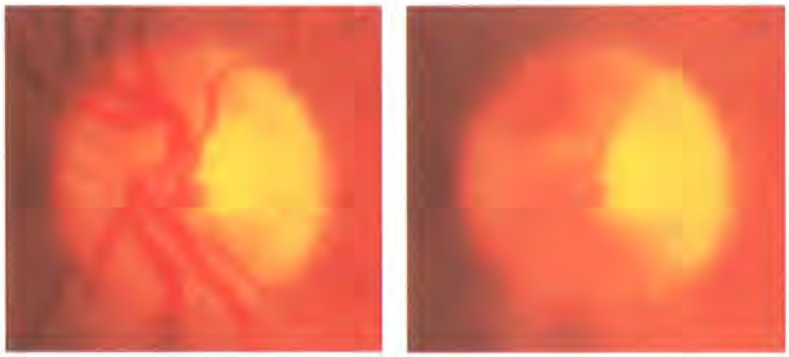


Figure 3.7: The Optic Nerve Head. a) Normal. b) Morphologically Closed.

These papers indicate the significant difficulties in segmenting the optic nerve head rim. However, neither author fully exploited the strongly consistent overall shape of the optic nerve head to constrain the contour, bypass distractors, and maintain shape where the rim is not visible. Both authors assumed that the model position was initialized quite accurately, as in their images the optic nerve head is always centered. In this research the problem domain is less constrained and initialisation is less reliable, with fundus images roughly centred on the fovea instead of the optic disc.

Hu *et al.* [40] described a method for boundary extraction of cross-sectional blood vessels in 3D imaging. A circular global model is first fit to the boundary contour, then a local deformable model with variable stiffness is closely fitted to strong edge features, while maintaining a smooth contour close to the global model when edge features are weak or missing. This next section extends on the work of Hu *et al.* for optic nerve head segmentation, making alterations as necessary, including generic improvements (elliptical model, use of the gradient vector, and fast optimization), and domain specific improvements (exploitation of the optic nerve head topology).

Hu's Circular Deformable Model

This section describes Hu's deformable model [40], with some change of variables from the original paper to simplify the presentation. Hu's deformable model has two stages; global and local. During the global stage, the circular model allows positional and radius changes. Under the control of a stiffness factor, the local level stage deformation from the global model, enabling a more accurate

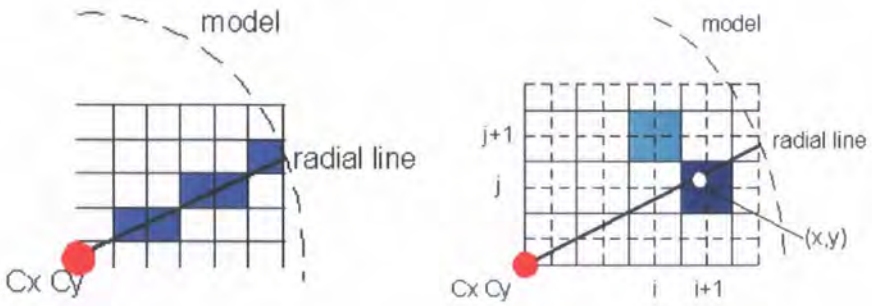


Figure 3.8: a) Irregular image sampling. b) Bilinear interpolation of gradient.

fit to the boundary.

The deformable model is defined by the center $\mathbf{c}$, S evenly-spaced radial spokes, each with angle θ_i, and direction vector $\mathbf{s}_i = [\cos(\theta_i), \sin(\theta_i)]$. The distances m_i from $\mathbf{c}$ along each spoke with corresponding (x, y) location vector is defined by $\mathbf{m}_i = \mathbf{c} + m_i\mathbf{s}_i$. In its global form, the model is a circle with radius $r = \overline{m}_i$, the local model's mean radial displacement. Radial spokes sample normalized gradient magnitude images, $\Gamma = \|\nabla\mathbf{I}\| / \max(\|\nabla\mathbf{I}\|)$ to produce radial gradient profiles. For each spoke $[i..S]$, the radial gradient profile is searched within a limited range about m_i, for high contrast changes; the "peak gradient" denoted by g_i, provides an external attractor point. The derivatives along the radial direction from inside to outside are computed. The sign of the derivatives are used to weight the normalised gradient; a positive sign represents dark to bright edge in the radial direction and a minus sign means a bright to dark edge. Since radial lines cross image pixels in an irregular manner, non-uniform sampling can occur; see figure 3.8a.

The gradient magnitude at distance ρ along spoke i is determined as follows. Let $\mathbf{p} = (p_x, p_y) = \mathbf{c} + \rho\mathbf{s}_i$. The radial gradient $\gamma_i(\rho)$ is calculated by bilinear interpolation of the image gradient magnitude at the point's four neighboring pixels [41]; see equation 3.1, figure 3.8b. The "peak," g_i, is chosen so that $\gamma_i(g_i)$ is the minimum on the spoke.

$$\gamma_i(\rho) = -\sum_{x,y} w_{x,y}\Gamma(x,y) \tag{3.1}$$

$$x_d = |p_x - x| \quad y_d = |p_y - y| \tag{3.2}$$

$$w_{x,y} = \begin{cases} (1-x_d)(1-y_d) & (x_d < 1) \wedge (y_d < 1) \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

The model's contour is altered under the influence of a force, f_i, with external and internal components; see equation 3.4. All the forces work along the radial spokes. The external force draws the model towards the attractor points of strong edge features, encouraging model deformation; see equation 3.5. The internal force comprises of two components that oppose the external force by limiting model deformation: the global force, which pulls the model towards the global shape; and the local smoothing force, which smoothes the model by penalizing sporadic deformation between neighboring spokes. The internal force which regulates model deformation is itself regulated by a stiffness factor, β_i, described further below.

$$f_i = f_i^{ext} - f_i^{int} \tag{3.4}$$

$$f_i^{ext} = g_i - m_i \tag{3.5}$$

$$f_i^{int} = \beta_i \left(d_i^{int} - \alpha \langle d_i^{int} \rangle \right) \tag{3.6}$$

where $d_i^{int} = m_i - r$ is deformation of the model from the global model on spoke i, and $\langle d_i^{int} \rangle$ is the mean deformation of neighboring spokes. The coefficient α balances the local versus global internal forces.

The stiffness parameter, β_i, controls the relative strength of the internal and external forces. On a strong edge (high contrast), a small stiffness value is used, the external force dominates, and localization is not unduly compromised. Where the edge is weak (blurred or missing) a larger stiffness value is assigned, emphasizing the internal force, attracting a smoothed contour to the global model. Hu used a function shaped like that in figure 3.9a to determine the stiffness factor, but did not give the formula; the interpreted stiffness function β_i is defined by:

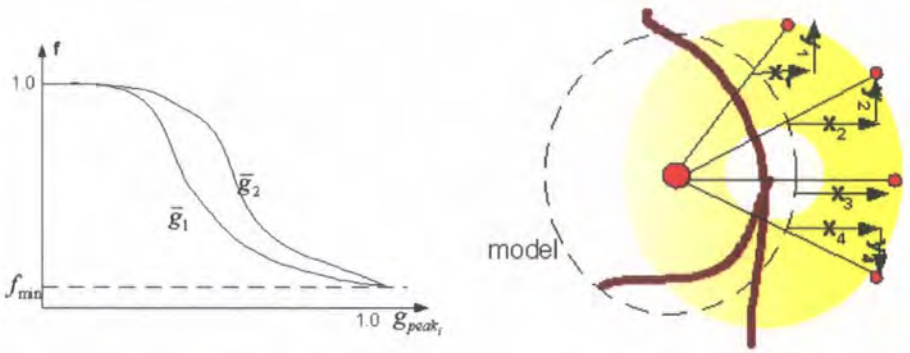


Figure 3.9: a) Stiffness factor function. b) Radial Force

$$\beta_i = \begin{cases} (2\gamma_i^3 - 3\gamma_i^2 + 1)^{\zeta} & \zeta \geq 0 \\ 1 - (2\psi_i^3 - 3\psi_i^2 + 1)^{(1-\zeta)} & \zeta < 0 \end{cases} \tag{3.7}$$

where γ_i represents the gradient magnitude at point g_i, (see equation 3.1), the mean gradient magnitude of the neighboring spokes is denoted by $\langle \gamma_i \rangle$. The squashing factor which shifts the function to the left or right of the middle position is defined by $\zeta = \chi + \langle \gamma_i \rangle$ where $\chi = 1.5$, and $\psi_i = 1 - \gamma_i$.

Hu uses the radial forces defined above (equation 3.4, 3.5, 3.6 and 3.7) to provide a center-shifting force, a global radius force, and local deformation forces (see figure 3.9b). The center-shifting force is found by summing and converting the radial forces exerted on each spoke into vector forces in the (x, y) plane (equation 3.8). The global radius force is found by subtracting the radial shifting force from the radial forces; see equation 3.10.

$$\mathbf{f} = \sum_{i=1}^{S} f_i \mathbf{s}_i \tag{3.8}$$

$$\overline{f_i^*} = \frac{1}{S} \sum_{i=1}^{S} f_i^* \tag{3.9}$$

$$f_i^* = f_i - \mathbf{s}_i . \bar{\mathbf{f}} \tag{3.10}$$

$$\bar{\mathbf{f}} = \frac{1}{S} \sum_{i=1}^{S} f_i \mathbf{s}_i \tag{3.11}$$

These forces are used to iteratively adjust the model; during the model's initial global stage, the

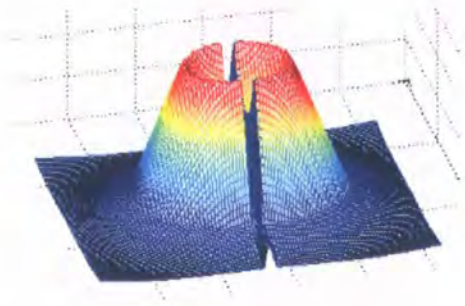


Figure 3.10: Localisation filter

global shifting force $\mathbf{f}$ is added to the model's centre $\mathbf{c}$, and the global radial force $\overline{f_i^*}$ is added to the radius. Once the model has reached equilibrium with the external force, the model decouples from its global form, allowing local deformation. During the model's local stage, the radial forces f_i^* are are added to deformation point m_i for radial spokes $[i..S]$.

3.3 Method

3.3.1 Optic Nerve Head Localisation

The localization algorithm uses a specialized correlation filter, which matches key elements of the optic disc structure. The correlation peak gives the approximate center of the optic disc. The optic disc consists of a high intensity near-circular rim, with a significant vertically-oriented, roughly centrally located band of low intensity blood vessels; other parts of the disc (including the interior) are not reliable and are discounted. The template consists of a Laplacian of Gaussian with a vertical channel in the middle to correspond to the major vessel band; see figure 3.10.

The size of the optic nerve head varies significantly; data set widths varied from 65–101 pixels (mean 78.5, standard deviation 7.6). This might suggest a need for filters at different scales. In fact the single filter used suffices; where the size mismatches the maximal correlation lies on an annulus around the optic nerve head center; the peak is therefore still located within the optic nerve head, which is sufficient for our requirements.

The template is correlated with the intensity component of the fundus image. Full Pearson-R correlation is used to account for variations in mean intensity and contrast, defined by:

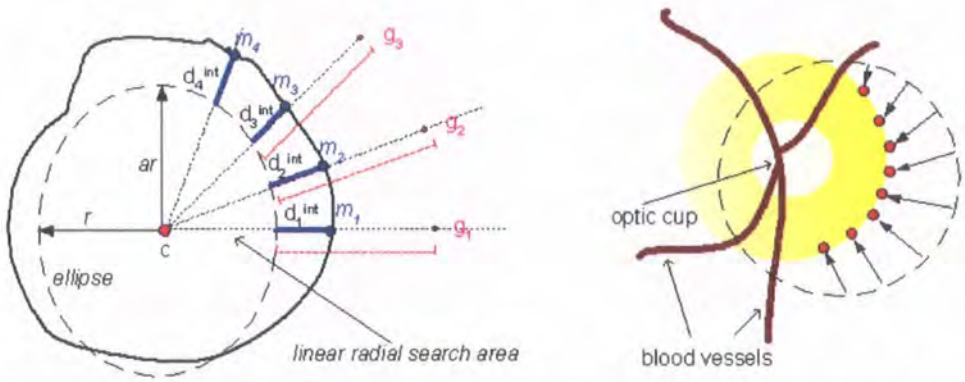


Figure 3.11: Model fitting. a) The deformable model (based on Hu *et al.*'s model) b) Locking to the temporal edge

$$C_{i,j} = \frac{\sum_{x,y}(f(x,y) - \overline{f}(x,y))(w(x-i,y-j) - \overline{w})}{\sum_{x,y}(f(x,y) - \overline{f}(x,y))^2 \sum_{x,y}(w(x-i,y-j) - \overline{w})^2} \tag{3.12}$$

where $\overline{w}$ is the mean value of the template, calculated once, and $\overline{f}$ is the mean value of the area covered by w. The filter is prone to locate a point slightly to the temporal side, due to the characteristic asymmetry of the optic nerve head (see figure 3.2), which is convenient for the next stage of the algorithm.

3.3.2 Optic Nerve Head Boundary Segmentation - Novel Alterations to the Hu's model

Three modifications to Hu *et al.*'s algorithm are presented: an elliptical global model; the use of the vector gradient; and the use of energy functions to support fast non-linear optimization. The global model is defined as an ellipse with a vertical principal axis, and a fixed aspect ratio, a; see figure 3.11a. The ellipse width is denoted by the "radius" parameter, r; and the height is ar. By scaling the elliptical model with the spoke ratios, a_i (equation 3.13), radial distances and forces are "normalized", equivalent to a circular model. The radius of the global model is defined by summing the model distance m_i from the model center $\mathbf{c}$, divided by aspect ratio a_i; see equation 3.14. The variables for the model and attractor points normalized to the circular frame are expressed by: $m_i^n = m_i/a_i$, $g_i^n = g_i/a_i$.

$$a_i = \|[\cos(\theta_i), a\sin(\theta_i)]\| \tag{3.13}$$

$$r = \frac{1}{S}\sum_i \frac{m_i}{a_i} \tag{3.14}$$

Two main factors affect the performance of active contours; the presence of noise, which can trap a model in a local minima; and the domain attraction of relevant edges. A smoothed normalized gradient vector image, $\Upsilon = \nabla\mathbf{I}/\max(\|\nabla\mathbf{I}\|)$ (Hu *et al.* does not use smoothing), is used to smooth out noise and increase the attraction range of high gradient magnitudes.

During profiling, the gradient between the optic nerve head and the background retina is high to low. To exploit this direction-sensitive gradient and to ignore distractors at most vessel edges, the dot product of the gradient vector and a spoke direction vector is taken – see equation 3.15; the weights are determined by equation 3.2.

$$\gamma_i(\rho) = \left(\sum_{x,y} w_{x,y}\Upsilon(x,y)\right).\mathbf{s}_i \tag{3.15}$$

During local deformation, equilibrium is reached using a fast non-linear optimization procedure, Quasi-Newton [42]. In addition to a gradient function, this procedures requires an explicit energy function; defined as the sum-squared deformation of the model from the attractors. The energy functions of the center-shifting force, total force, external force, global internal force, local internal force are represented by E^{xy}, E, E^{ext}, E^{glo}, E^{loc} respectively:

$$E^{xy} = \frac{1}{2}\left\|\left(\frac{1}{S}\sum_i \mathbf{m_i^n} - \mathbf{c}\right)\right\|^2 \tag{3.16}$$

$$E = E^{ext} + E^{int} = E^{ext} + E^{glo} + E^{loc} \tag{3.17}$$

$$E^{ext} = \frac{1}{2}\sum_i (g_i^n - m_i^n)^2 \tag{3.18}$$

$$E^{glo} = \frac{1}{2}\sum_i \beta_i (m_i^n - r)^2 \tag{3.19}$$

$$E^{loc} = \frac{1}{2}\sum_i \beta_i \alpha (m_i^n - \langle m_i^n \rangle)^2 \tag{3.20}$$

Differentiating E^{xy} with respect to $\mathbf{c}$, and E with respect to m_i^n, and observing that r and $\langle m_i^n \rangle$ in equations 3.19 and 3.20 are both dependent on m_i^n, the following gradients are obtained:

$$\frac{dE^{xy}}{d\mathbf{c}} = \frac{1}{S}\sum_i m_i^n \mathbf{s_i} \tag{3.21}$$

$$\frac{dE}{dm_i^n} = \frac{dE^{ext}}{dm_i^n} + \frac{dE^{glo}}{dm_i^n} + \frac{dE^{loc}}{dm_i^n} \tag{3.22}$$

$$\frac{dE^{ext}}{dm_i^n} = g_i^n - m_i^n \tag{3.23}$$

$$\frac{dE^{glo}}{dm_i^n} = \beta_i((m_i^n - \overline{m_i^n}) - \overline{(m_i^n - \overline{m_i^n})}) \tag{3.24}$$

$$\frac{dE^{loc}}{dm_i^n} = \beta_i\alpha\left[(m_i^n - \langle m_i^n \rangle) - \langle [m_i^n - \langle m_i^n \rangle] \rangle\right] \tag{3.25}$$

Let $\mathbf{w}$ denote a vector of model parameters, $E(\mathbf{w})$ the error function for these parameters, and $\mathbf{w}'$ the gradient of the error function with respect to the parameters; $\mathbf{w}$ may be any desired combination of model parameters (e.g. m_i, $\mathbf{c}$, r); if we wish to optimize several parameters simultaneously, the appropriate parameter and gradient vectors may be concatenated and the error functions summed. The global and local model use different optimisation techniques: gradient descent with momentum (for the simpler global model phase), and Quasi-Newton BFGS (for the local model phase); both proved more efficient than the direct iterative technique suggested by Hu *et al.*. In gradient descent with momentum, a learning rate, η, and a momentum rate, μ, are selected; the model parameters $\mathbf{w}$ are iteratively updated at $\tau + 1$ using (the number of iterations required do not justify a more sophisticated minimisation procedure):

$$\mathbf{w}_{\tau+1} = \eta \mathbf{w'}_\tau + \mu \mathbf{w}_\tau \tag{3.26}$$

Quasi-Newton optimization involves generating a sequence of matrices, $\mathbf{G}_\tau$ representing increasing accurate approximations to the inverse Hessian, $\mathbf{H}^{-1}$, using information from the first derivatives of the error function [43]. Using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) procedure, the approximation $\mathbf{G}_\tau$ of the inverse Hessian matrix, is updated using equation 3.31.

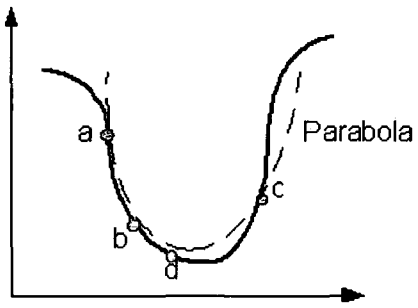


Figure 3.12: Line Search

$$\mathbf{G}_0 = -\mathbf{I}\mathbf{w'}_0 \tag{3.27}$$

$$\mathbf{p} = \mathbf{w}_{\tau+1} - \mathbf{w}_\tau \tag{3.28}$$

$$\nu = \mathbf{w'}_{\tau+1} - \mathbf{w'}_\tau \tag{3.29}$$

$$\mathbf{u} = \frac{\mathbf{p}}{\mathbf{p}^T\nu} - \frac{\mathbf{G}_\tau{}^T\nu}{\nu^T\mathbf{G}_\tau{}^T\nu} \tag{3.30}$$

$$\mathbf{G}_{\tau+1} = \mathbf{G}_\tau + \frac{\mathbf{p}\mathbf{G}_\tau{}^T}{\mathbf{G}_\tau{}^T\nu} - \frac{(\mathbf{G}_\tau\nu)\nu^T\mathbf{G}_\tau}{\nu^T\mathbf{G}_\tau\nu} + (\nu^T\mathbf{G}_\tau\nu)\mathbf{u}\mathbf{u}^T \tag{3.31}$$

$$\mathbf{w}_{(\tau+1)} = \mathbf{w}_{(\tau)} + \xi_\tau\mathbf{G}\mathbf{w'} \tag{3.32}$$

The update procedure ensures that the approximation to the inverse Hessian is positive definite. The direction vector $-\mathbf{G}_\tau\mathbf{w'}$ is guaranteed to descend, and rapidly converges on the Newton direction. Optimization is by equation 3.32, where ξ_τ is found by a bracketing line search.

Line searching represents a one-dimensional minimisation problem, consisting of two stages. In the first phase, the directional line is searched for three points a, b and c that brackets the minimum, such that a>b>c. The second phase is to find the minimum. This is achieved by fitting a quadratic polynomial to the error function evaluating at points a, b, and c, moving to the minimum of the parabola at point d (see figure 3.12). This process is repeated, calculating the error function at the new point and fitting a new parabola to the three points in this case b, d, c where b>d<c. The line search is terminated when a satisfactory solution is found [42].

3.3.3 Optic Nerve Head Boundary Segmentation - Phases

The presented localisation and segmentation algorithms are very reliable, although the latter is somewhat sensitive to poor initialization. Optimization of the global model is the most critical phase of this algorithm; however, direct fitting of this proves unreliable. A four phase algorithm is introduced, with the global fitting carried across two phases. Phase one (localisation) was described in section 3.3.1. In phase two, the global model roughly locates the rim, by locking the model's temporal edge onto the relatively reliable and distractor-free temporal rim. Phase three fits the model to the entire rim. Phase four decouples from the global model, allowing deformation of the local model to the optic boundary; phases two to four are described below.

Phase two – temporal lock

In many fundal images the temporal side of the optic nerve head has a stronger edge than its nasal counterpart; which often has a weak or non existent edge. A weak nasal edge can be problematic to the nasal side of the model, often resulting in convergence on the blood vessels or the pallor. This problem is tackled by initially restricting the model's active spokes to the temporal side.

The temporal lock is defined as a global model using $S = 9$ spokes at angles $-60°$ to $+60°$ in steps of $15°$, on the temporal edge only. The attraction range of the temporal edge is increased by smoothing the gradient image with a gaussian filter ($\sigma = 5.0$), and the radial search range for γ_i is set to $m_i \pm 6$ pixels. The aspect ratio is fixed at $a = 1.03$ and the radius is set to the average optic disc radius $r = 40$; only the model center is adjusted. Optimization is by gradient descent (see equation 3.26), with $\eta = 1$, $\mu = 0.1$, $\mathbf{w} = \mathbf{c}$, $\mathbf{w}' = dE^{xy}/d\mathbf{c}$ (see equation 3.21).

The temporal bias of the localization algorithm ensures that the initial contour is usually just outside the temporal edge. Due to the lack of temporal vessels, their typically radial alignment, and the relative strength of the temporal edge, this phase locks on with high reliability; see figure 3.11b. The fixed radius ensures that the nasal edge of the model bypasses the distractors at the pallor and central blood vessels. A fixed radius is necessary since fitting an elliptical model to $120°$ of arc on one side is prone to gross mis-estimation.

Phase three – global fit

Once the temporally active model reaches equilibrium with the temporal optic edge, the remaining spokes are activated, using $S = 24$ evenly-spaced spokes at 15° spacing; the radial search range is reduced to ±4 pixels. Other parameters are identical to phase two.

In addition to the optic center, the radius of the model is now allowed to alter; the energy function becomes $E^{xy} + E^{ext}$, with the latter treated as a function of r; $\mathbf{w} = [\mathbf{c}; r]$. The gradient of E^{ext} with respect to r is given by equation 3.33, derived from equation 3.23, the global model is enforced by setting $m_i = a_i r$.

$$\frac{dE^{ext}}{dr} = \sum_i \frac{dE^{ext}}{dm_i^n} \frac{dm_i^n}{dr} = \sum_i (g_i - r) \tag{3.33}$$

The model is "tweaked" onto the nasal edge using short burst of gradient descent. Using the gradient vector (equation 3.15), the model avoids being pulled onto the nasal edges of the nasal vessel, or the edges of near-orthogonal vertically-departing vessels, due to their opposite contrast direction to the rim. The pallor edge and peripapillary artifacts are the only strong boundary distractors. The small search range usually avoids the former.

The fixed aspect ratio, $a = 1.03$ may lead to sub-optimal models (actual aspect ratios were in the range 0.94–1.15, mean 1.03, s.d. 0.038). To compensate for this, the aspect ratio from the attractor points are iteratively recalculated at the end of the phase, and the optimization is repeated (which is extremely fast); three iterations suffice. The aspect ratio is calculated using equation 3.34, where $\mathbf{u}$ and $\mathbf{v}$ are the unit vectors along the x and y axes.

$$a = \sqrt{\frac{\sum_i m_i (v.\mathbf{s_i})^2}{\sum_i m_i (\mathbf{u}.s_i)^2}} \tag{3.34}$$

Phase four – local deformation

When the global model reaches equilibrium the local model is activated, allowing the contour to deform from its elliptical shape under the influence of the local edge features but still under the restraint of the global model. To improve contour localization the gradient image is recalculated with smoothing factor $\sigma = 1$. This reduces the attraction of the optic edge but allows the model contour to get closer to the actual optic boundary. The model points, m_i, are allowed to vary

in addition to the center, $\mathbf{c}$; therefore, $\mathbf{w} = [\mathbf{c}; m_1, \ldots, m_S]$. The model is allowed to globally reposition, in order to fine tune it's central location. The full energy function from equation 3.17 is used, with $\mathbf{w}'$ composed from the corresponding gradients. As the local optimization stage is more challenging than the global stage, a faster optimization algorithm is used; Quasi-Newton. The control parameter settings for the energy function were, $\alpha = 0.5$.

3.4 Evaluation

Performance of the optic nerve head location and segmentation algorithms are evaluated against a random sample of 100 fundal images taken from 50 patients attending the diabetic retinal-screening programme at City Hospital, Birmingham. The images were acquired using a Canon CR6 45MNf fundus camera, with a field angle lens of 45 degrees, resolution 640 × 480. There is considerable variation in the images, with many characteristics that can affect the algorithms. Image distractors that present potential problems are summarized in Table 3.1. Some images have multiple characteristics. Distractors are divided into three categories; potentially ungradable, affecting localization; and affecting segmentation.

Table 3.1: Screening Data Image Quality

Characteristic	No. images
No detectable optic nerve head	4
Severe Cataract	8
Moderate Cataract	2
Total potentially unusable	10
Exudates or laser scars	7
Light artifacts	7
Easily visible choroidal vessels	20
Total localization endangered	34
Some of rim blurred or missing	27
Severe peripapillary atrophy	6
Moderate peripapillary atrophy	19
Concentric peripapillary atrophy/artifacts	23
Strong pallor distractor	13
Total segmentation endangered	58

3.4.1 Optic Nerve Head Localization Results

Performance is assessed against the 96 out of our 100 fundus images that have a discernable optic disc. An ophthalmologist labelled the center of each optic disc to create a gold standard. Images are graded by distance from the true optic center (see table 3.2); the algorithm usually finds the center within 15 pixels, which is acceptable as a starting point for rim segmentation. The center point is within the optic disc area on all usable images bar one. Figure 3.13 shows some examples of its performance. This simple approach yields excellent results and is robust; its success lies in careful design of the filter.

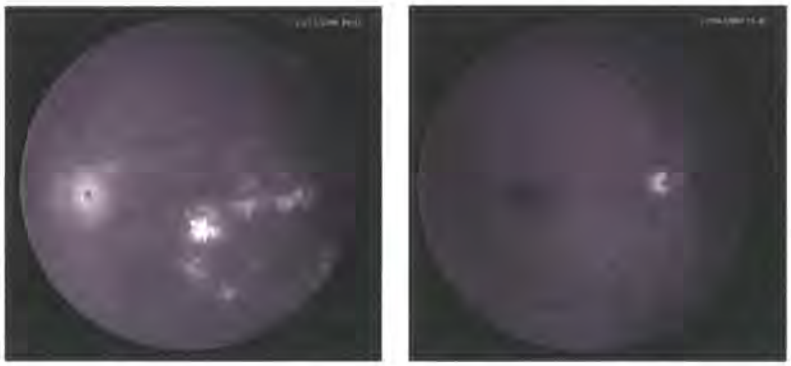


Figure 3.13: Localization Algorithm. a) Gross exudation. b) Strong pallor.

Table 3.2: Localization Performance

Pixel Distance	0 – 5	6 – 15	16 – 25	26+
Fundus images	42	45	8	1
Percentage	43.8	46.9	8.3	1

3.4.2 Optic Nerve Head Boundary Segmentation Results

The segmentation algorithm is tested against a number of alternative approaches, described below. Performance was evaluated against a subset of 90 images, excluding those with no discernable optic disc, or with severe enough cataract to prevent meaningful segmentation. A "gold standard" segmentation of the optic boundary was produced by four clinicians manually delimiting the optic rim. Their mean contours, and the radial standard deviations of these contours were calculated. Let μ_i^j and σ_i^j denote the clinician's rim location on spoke i of image j. The discrepancy δ^j, on

image j, is defined by equation 3.35. Division by the standard deviation σ compensates for the clinicians uncertainty of rim position; where the clinicians are in exact agreement, a small factor $\varepsilon = 0.5$ prevents division by zero.

$$\delta^j = \sum_i \frac{|m_i^j - \mu_i^j|}{\sigma_i^j + \varepsilon} \tag{3.35}$$

For illustrative purposes the ogive (cumulative frequency polygon) of disparity, δ using a logarithmic scale (i.e. the number of images with disparity less than δ, $|\{j : \delta^j < \delta\}|$, versus δ) is plotted. Within these plots, the superior model lies left of and above an inferior one; the number of images fit to any given level of accuracy can be read off the y-axis.

Parameter Settings

The parameter values reported in this chapter were heuristically selected for best performance - a wide range of parameter settings were appraised for: initial model diameter and aspect ratio, gaussian smoothing factors, radial profile search sizes, and α in the deformable model. Generally, the algorithm's overall performance is relatively insensitive to changes in these parameter values.

The exception, however, is the initial diameter parameter which has significant performance implications for a small number of of outlying sized optic nerves heads, as pallor and peripapillary distractors may induce errors. This might suggest running the algorithm with several initial diameters and selecting the lowest energy resulting contour. Unfortunately, two problems exist with this approach. First, the strong distractors of the pallor and peripapillary often having a lower energy than the true contour; second, with pallor distractors being inside the true rim, and peripapillary distractors being outside, a simple geometric rule could not choose between alternative solutions (which makes the utility of even sophisticated minima-avoidance algorithms [44] questionable). The initial diameter is therefore fixed which functions well in the vast majority of cases.

Performance of the algorithm

Figure 3.15a shows the performance graph of the final algorithm (using both temporal lock and vector gradient) against a simple benchmark approach ("direct"). Qualitatively, four categories are defined (Excellent, Good, Fair and Poor) containing images with benchmark disparity up to

one, two, five, or more, respectively. These four disparity ranges correspond well to the subjective assessment of performance quality. The performance on this subjective scale is summarised in Table 3.3 and examples within each category shown in Figure 3.14.

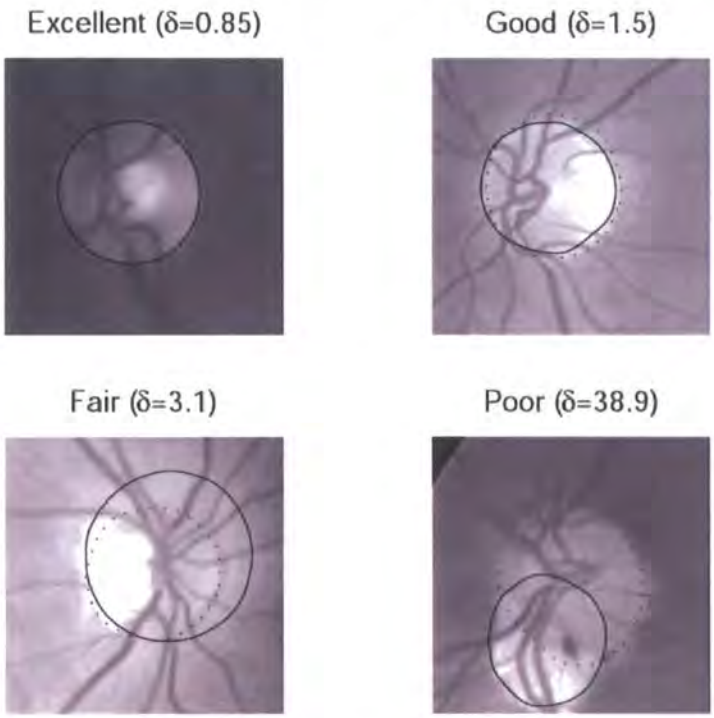


Figure 3.14: Sample Segmentations. a) Excellent. b) Good. c) Fair. d) Poor. Solid line: algorithm; dotted line: mean clinician boundary.

Table 3.3: Subjective Classification of Performance

(%)	*Excellent*	*Good*	*Fair*	*Poor*
TemporalLock	42	31	10	17
Direct	9	8	30	53
$DV - Hough$	39	22	20	19

Comparison with Temporal Lock Alternatives

The temporal lock phase is a critical aspect of the algorithm that compensates for poor localisation while avoiding distractors. To affirm the superiority of the temporal lock, performance is measured against alternative methods that attempt to remove or mitigate distractors and improve localization

of the rim. The removal of optic blood vessels by morphology is a method suggested by both Lee [32] and Mendels [39]. This is referred to as "de-vascularization", and is implemented by grey-scale morphological closure with a disc-shaped structuring element of radius 7. The effectiveness of the vector gradient over the magnitude is shown in figure 3.15b, together with the interaction between de-vascularization. Mendels noted that the use of the vector gradient rather than magnitude improved his algorithm, however there is no advantage in using both techniques as they compensate for the same distractors.

Fitting an elliptical Hough transform to the optic rim is an non-iterative alternative to the temporal lock phase. In this comparison the image is processed using a Canny edge-detector (upper and lower thresholds 0.2 and 0.08 respectively). Due to the vessel edges distractors, it was found that the Hough transform only worked well in conjunction with de-vascularization. Ellipses were of width 56–112, aspect ratio 1.03; the Hough-space was smoothed with a Gaussian kernel ($\sigma = 1.5$) before peak detection to account for deformations. It was found the Hough technique was too unreliable when implemented across the entire image, so the Hough filter was used to tune the location within a search radius of 150 pixels initialised by the phase one localization algorithm; the Hough center was only accepted if the peak in Hough space exceeded a threshold.

In combination with de-vascularization and the vector gradient, an algorithm ("DV-Hough") emerged with performance close to that of the temporal lock algorithm, as illustrated on figure 3.15c. Using the one-sided sign test for paired medians, the difference in performance (temporal lock is better in 59/90 cases) is significant at the 0.5% level. The superiority of "DV-Hough" and "temporal lock" algorithm over the simpler alternatives is clear-cut. Integration of the temporal lock algorithm with de-vascularization and the Hough search were tried, but these additions did not improve that algorithm.

Although the temporal lock algorithm performs well in comparison to alternatives, there is room for improvement; the algorithm fails to detect the rim in a number of cases where it is unambiguously apparent to the human observer. The attributed cause of failure on 15 "poor" performance images are shown in Table 3.4; two of the "temporal lock" failures are difficult to attribute cleanly, as localization was rather poor and probably contributed to the failure. Two of the "peripapillary atrophy" failures subjectively look "fair."

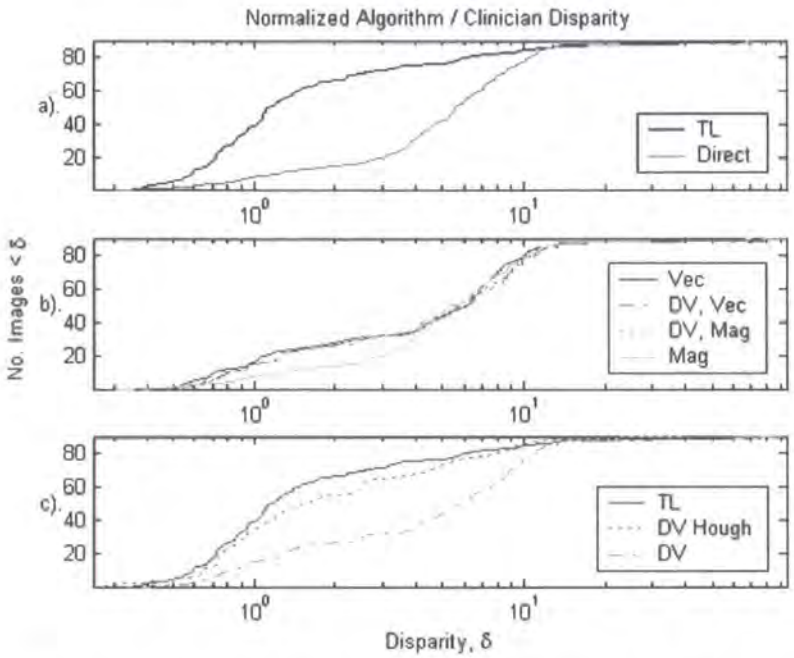


Figure 3.15: a). Temporal Lock (TL) versus direct algorithm. b). Interaction between De-vascularization (DV) and vector (Vec)/magnitude (Mag) gradient versions of direct algorithm. c). Temporal Lock (TL), De-vascularized Hough (DV Hough), and De-vascularized direct (DV).

3.5 Conclusion

Algorithms for the localisation and segmentation of the optic nerve head have been presented; an important stage in structured analysis of the retina. Although a number of methods have been published for optic nerve head localisation, many are unreliable when confronted with images of diseased retinae including strong distractors (such as circinate exudates), and the reliable methods tend to be quite computationally complex. A simple but effective algorithm for localisation has been presented.

Table 3.4: Error types

Source of error	Number
Localization failure	3
Temporal lock failure	5
Pallor distractor	3
Severe peripapillary atrophy	4

Optic nerve head segmentation by active contours has not been extensively examined in the past. There are significant problems in dealing with distractors along blood vessels edges and the pallor, and with the very variable appearance of the optic nerve head. Previously published techniques require careful initialisation of the model position, pre-processing of the image using morphological operations, and perform badly where the rim is faint or undetectable. In contrast, the algorithm presented in this chapter exploits specific features of the optic nerve head anatomy to achieve good localization while avoiding distractors. The temporal lock algorithm exploits the natural shape of the neuroretinal rim to bypass blood vessels and avoid the pallor, and the global and local deformable model deals effectively with weak areas of rim and vessel crossings. Additionally, applying a Gaussian smoothing operator to the pre-processed image strengthens the optic disc's contour. In the global optimisation stage, the inclusion of a momentum term and learning rate stabilizes the movement and growth of the active contour, and decreases optimisation time. Defining energy functions and using a Quasi-Newton optimization strategy makes the local phase of the algorithm reasonably fast.

Experimental comparisons have been conducted with a range of alternative approaches, using a randomly-selected experimental set, and have demonstrated the superiority of the proposed approach. The imperfect boundary locations are largely due to the variable nature of the images and the presence of distractor boundaries concentric with the desired rim, which may be located either inside or outside the rim. To improve this algorithm, future work would require dealing with these distractors and would require a higher level of reasoning.

Although the current approach focuses on detecting the optic nerve head boundary as a stage in the structured analysis of images for diabetic retinopathy fundus images, the approach is also of relevance in diagnosis of other diseases, and particularly in screening for glaucoma.

Chapter 4

Fovea Detection & Vascular Segmentation

4.1 Introduction

This chapter presents two existing algorithms for the detection and segmentation of the fovea and the vascular network. For completeness, Sinthanayothin's *et al.* [25] fovea detection algorithm and Hunter's *et al.* [45] vascular segmentation algorithm are described in detail.

Due to the eloquent simplicity and segmentation performance of their respective approaches, no further investigation into other approaches were sought and their work is incorporated into the overall retinal analysis system.

4.1.1 Chapter Contents

In section 4.2 the importance and difficulty of foveal detection is described together with a detailed description of Sinthanayothin's algorithm. The performance of the algorithm is evaluated against 1000 screening images. In section 4.3, the motivation, appearance and distractors of vascular segmentation are explained. This is followed by a description of Hunter's algorithm. The performance of Hunter's technique concludes this chapter.

4.2 Fovea Detection

4.2.1 Motivation

The fovea is a vital part of the retina responsible for central, sharp colour vision. As diabetic retinopathy progresses, retinal lesions caused by vascular leakage can obscure areas of the retina. If these lesions occur near the fovea, there is an increased risk of foveal obscurement and loss of central vision. The macula is defined as one optic disc diameter around the fovea. Any retinopathy within this area is known as diabetic maculopathy. Identifying the centre of the fovea is therefore vital to enable accurate diagnoses of diabetic maculopathy.

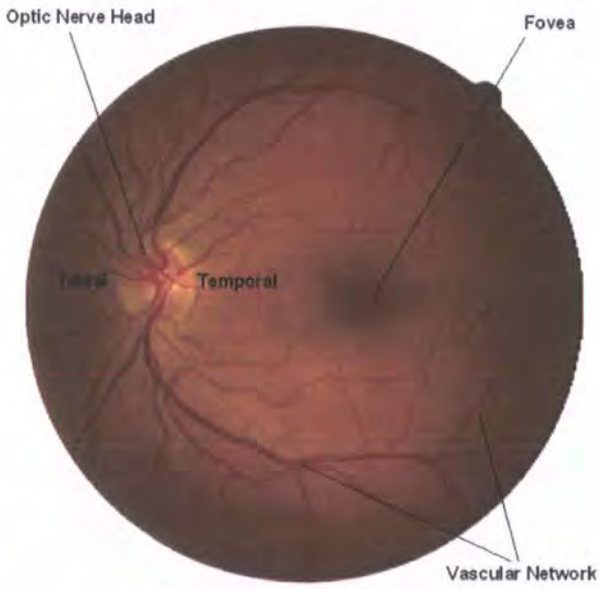


Figure 4.1: Fovea

4.2.2 Appearance and Distractors

The fovea is situated in the centre of the fundus and approximately 2.5 disk diameters temporal to the temporal edge of the optic nerve head; see figure 4.1. The fovea is typically the darkest region of the fundus with a similar intensity to blood vessels. A circular indentation in the retina and 1.5mm in diameter [12], the fovea is approximately half the diameter of the optic disc (fovea diameter 40 pixels in a 760×570 image). Unlike the peripheral retina, the fovea has no blood

vessels and commonly has low contrast. An example of a fovea with very low contrast is shown in figure 4.2a. The level of foveal contrast is a serious issue faced by location algorithms, especially when combined with higher contrast features within the proximity of the real fovea.

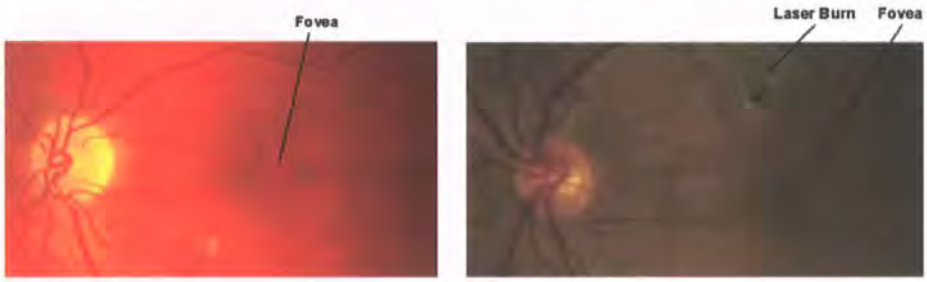


Figure 4.2: Fovea. a) Low Contrast. b) Distracter.

Large circular haemorrhages and (dark) laser burn clusters are the main form of distractor for foveal location algorithms; these features often having a higher contrast with the background retina than the fovea. This is depicted in figure 4.2b. where a laser burn has greater contrast than the fovea.

4.2.3 Fovea Location Algorithm

Sinthanayothin *et al.* [25] present a fovea location algorithm that is based on a simple template matching technique - a template approximating a typical fovea is convolved over an intensity fundus image, producing a correlation image showing the degree of intensity similarity between the template and fundus at each pixel location. The maximum correlation between the template and the intensity image, is deemed the centre of the fovea.

The foveal template as depicted in figure 4.3 is based on a Gaussian distribution of size 40×40 pixels and standard distribution of $\sigma = 22$. The template is defined by equation 4.1.

$$g(i,j) = 128\left[1 - \frac{1}{2}exp\left(\frac{-(i^2+j^2)}{2\sigma^2}\right)\right] \tag{4.1}$$

where (i,j) represents the centre of the template. The foveal template is correlated with the intensity component of the fundus image. A subimage $W(i,j)$ of dimensions $M \times M$ is centred on pixel (i,j) with template intensities $g(k,l)$, $(k,l) \in W(i,j)$. The correlation coefficient of W at (i,j) with image intensities of $f(i,j)$ is defined by equation 4.2

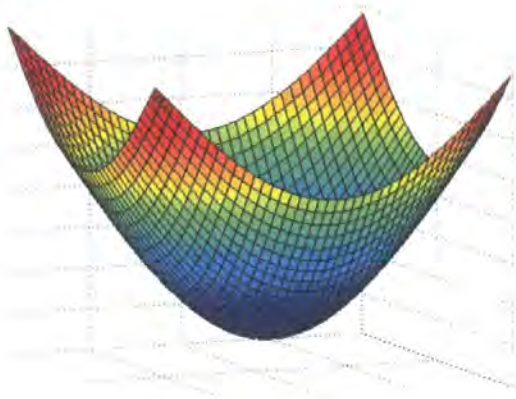


Figure 4.3: Fovea Template

$$\gamma(i,j) = \frac{\sum_{k,l}[f(k,l) - < f >_w][g(i-k, j-1) - < g >_w]}{\left\{\sum_{k,l}(f(k,l) - < f >_w)^2 \sum_{k,l}(g(i-k, j-1) - < g >_w)^2\right\}^{\frac{1}{2}}} \tag{4.2}$$

In 71 foveal centred images, the algorithm is reported to correctly locate 60 fovea regions (84.5%), the other 11 (15.5%) were unidentified. An extensive evaluation of this algorithm is described below.

4.2.4 Evaluation

The accuracy and precision of Sinthanayothin's algorithm were tested using 914 gold standard foveal marked images obtained from 1000 fundal images; captured on a Canon CR6 45MNf fundus camera with a 45 degree field of view and resolution of 760×570. The foveal centre of all clinically gradable images were manually marked by an ophthalmologist. It is against this benchmark that the algorithm is evaluated. Out of the 1000 fundal images, the ophthalmologist excluded 86 images from study due to inferior image quality. The distance between the estimated and the benchmarked foveal centre is shown in table 4.1. As Sinthanayothin did not present the algorithm's accuracy, no comparison can be made with Sinthanayothin's own results.

Table 4.1 shows that 53.3% and 89.7% of the estimated fovea centres were within 10 and 20 pixels respectively of the benchmarked fovea centre. Generally, if the estimated fovea centre was

Table 4.1: Distance from Real Fovea Centre

Distance	Number of Images
0-5	202
6-10	291
11-20	336
21-25	56
26-30	24
31-40	11
41-50	1
50+	3
Image Total	924

Table 4.2: Accumulated Distance from Benchmarked Fovea

Template Size	Accumulated Distance
34	15492
36	13995
38	13440
40	12221
42	13202
44	12999
46	13391

greater than 25 pixels away from the benchmarked centre, the fovea was either extremely faint or obscured (by lesions or laser scars); making location by template matching difficult.

The average size of a fovea is 40 × 40 pixels, this is the same dimension as Sinthanayothin's foveal template. The size of template used by template matching techniques can significantly effect performance; a variety of fovea template sizes were evaluated by accumulating the distances between the estimated and benchmarked fovea; see table 4.2. Table 4.2 shows that the smallest accumulated distance and therefore the optimum template size is in fact 40 × 40 pixels as implemented by Sinthanayothin.

4.3 Vascular Segmentation

4.3.1 Motivation

Diabetic retinopathy is primarily a disease of blood vessels [45]. The change in width of retinal vessels within the fundus is believed to be indicative of the risk level of diabetic retinopathy [46]; venous beading (unusual variations in diameter along a vein) is thought to be one of the most

powerful predictors of proliferative diabetic retinopathy [47]. Another important severity indicator of diabetic retinopathy is neovascularization - new vessels that are tortuous in shape. Detecting both of these retinal complications requires the accurate identification of the vascular network. Vascular segmentation is also important in preventing false positive haemorrhage classifications where veins change angle within the retina or arteries cross causing small relative dark segments of vessel.

4.3.2 Appearance and Distractors

Within a fundus image, the diameter, tortuosity and contrast of blood vessels can vary significantly. Although blood vessels are usually darker than their surroundings, variable retinal pigmentation, uneven illumination and noise can cause varied contrast between the vascular network and background image. This can cause problems with linear segmentation filters that threshold images using a single value. The contrast between vessel and background may vary over an image, but the structure of blood vessel appearance remains constant - cross-vessel profiles tend to keep the same shape independent of contrast change (dark in the centre of the blood vessel and lighter either side of the blood vessel, see figure 4.5b).

In addition to uneven illumination and noise, vessel segmentation is complicated by a further two image distractors - locally dark channels and one side contrast boundaries. As shown in figure 4.4, locally dark channels can form in the spaces between exudate pairs, light artefacts (camera flash) and between visible choroidal vessels ("tiger stripes") and retinal background. One side contrast boundaries such as exudate edges can act as segmentation distracters even though a filter may respond to only one half of a light distractors edge - the response of the filter to that edge may be stronger than that of a blood vessel with lower overall contrast.

4.3.3 Vascular Segmentation Algorithm

Hunter *et al.* [45] exploits the structure of the blood vessels using a characteristic cross-vessel profile (see figure 4.5b). Consider three parallel lines 10 pixels in length, one within the vessel and two lying either side. If the three lines are sampled at the same orientation as a vessel, the centreline should be uniformly dark and the two outer lines uniformly light.

The contrast is measured between the central line and satellite "tram-lines" using a minimax

Figure 4.4: Distracters a) Exudates b) Light Artefact c) Tiger Strips

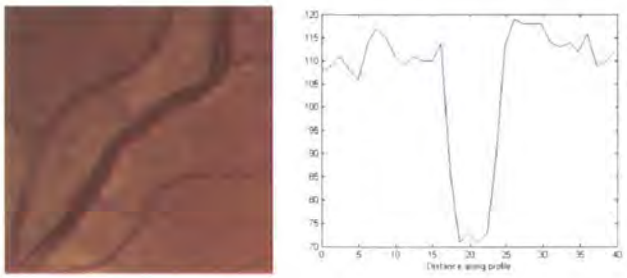


Figure 4.5: Normal a) Blood Vessel b) Blood Vessel Cross Section Profile

algorithm: the maximum (lightest intensity value) intensity of the inner line is subtracted from the minimum (darkest intensity value) intensity of the outer tram-lines. If the difference is positive then the interior tramline is darker than the outer tramlines and characteristic of a blood vessel. To desensitise the filter from the effects of noise, an order statistic filter is used to calculate the contrast: the inner and outer filters are sorted by intensity value, selecting the third darkest and third lightest intensity value respectively, then taking the difference. The tramline filter is applied using a number of rotational orientations, selecting the optimum response at each pixel. The orientation-sensitivity and order-statistic ensures the filter is reasonably robust against noise and insensitive to distractors, such as one-sided contrast boundaries found at the edge of bright exudates, light artifacts and the optic nerve head.

4.3.4 Evaluation

The tram-line filter's performance is measured against STARE's [48] well established vascular segmentation algorithm. Using STARE's publicly available test images and ground truth, the

tram-line algorithm proved superior to STARE when measured by vessel length, with 57.2% of the vascular network (by length) correctly identified with 2.9% false positives (excess length). This is compared with 57% and 8.5% for STARE. A blood vessel ground truth consists of manual tracings of the blood vessel network from a fundus image. The difficulty of providing a clinical blood vessel ground truth, has prevented the algorithm being evaluated on the 1000 Birmingham screening images (due to the limited clinician resources available to the project).

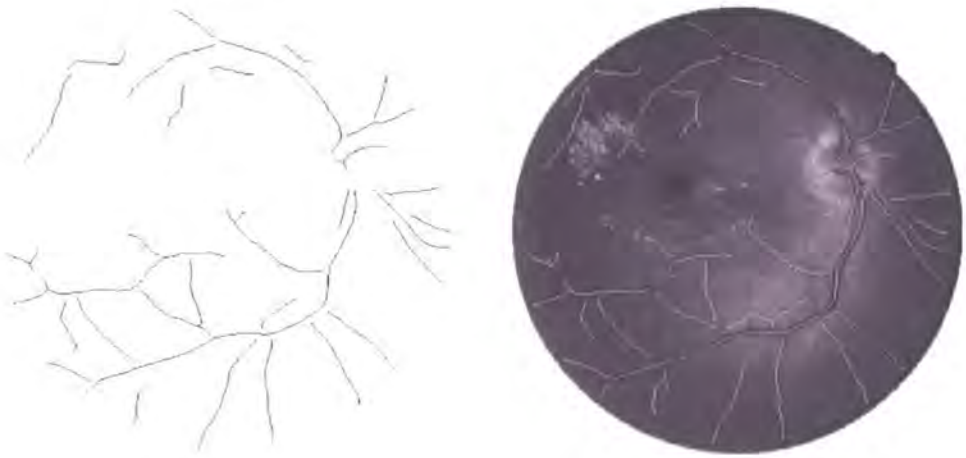


Figure 4.6:

The majority of unsegmented blood vessel are small capillaries that constantly come in and out of view. Figure 4.6 shows an example segmented vascular centreline and centreline fundus overlay. The additional research necessary to segment the remaining 42.8% of the vascular network is out of the scope of this research as lesion detection and classification has a higher priority.

Chapter 5

Diabetic Lesions

5.1 Introduction

This chapter presents an algorithm for the automated segmentation and classification of diabetic lesions in retinal images. Within this chapter, three types of lesions indicative of diabetic retinopathy (haemorrhages, microaneurysms and exudates) are located, segmented and classified. Potential lesions are identified using peak detection (creating seed points), with lesion boundaries defined using a novel region growing algorithm. In this algorithm regions are grown from seed points by sequentially adding adjoining pixels whose intensities are closest to the seed intensity. The contrast between the internal region and a dilated boundary is recorded at each iteration. Pixels are added in this way until the compactness of the region falls below a set threshold. The boundary with the highest contrast is selected to represent the potential lesion. Features relating to shape, size, and colour are extracted from the bounded regions and fed into an ensemble neural network. The region growth algorithm is evaluated against alternative approaches using a set of 100 lesion subimages taken from 20 down-sampled high definition fundus images. The results show that the algorithm is superior to previously published methods.

5.1.1 Motivation

The diagnosis of diabetic retinopathy is based upon visually recognizing various clinical features. Retinal lesions are among the first visual indicators suggestive of diabetic retinopathy. The threat

to visual loss increases with the frequency of retinal lesions combined with their encroachment into the macula (one optic disc diameter around the fovea). To enable early diagnosis, it is therefore necessary to identify both frequency and position of retinal lesions in relation to the fovea. Retinal lesions include haemorrhages, microaneurysms (dark spots); and exudates (bright spots). The appearance of microaneurysms is said to be the earliest indicator of diabetic retinopathy. This is commonly followed by the occurrence of haemorrhages and exudates which are characteristic of clinically significant retinopathy. Image distractors such as lighting variation, natural pigmentation, light artefacts, broken capillaries and non diabetic lesions make lesion identification and segmentation a challenging problem domain. In order to distinguish lesions from image distractors, sophisticated region descriptors and classification techniques are required. Image distractors affect bright and dark regions differently, consequently, lesion appearance, image distractors and segmentation literature are addressed for bright and dark lesions separately.

5.1.2 Chapter Contents

Section 5.2 is subdivided into three parts: bright lesions, dark lesions and classification. In the first two subsections the appearance of bright and dark diabetic lesions are described together with previously published lesion segmentation algorithms. In the third subsection image classification techniques and algorithms are discussed. In section 5.3, a new region growth algorithm is described together with region descriptors used in lesion classification. The new growth algorithm is evaluated against alternatives from the literature in section 5.4 and section 5.5 concludes the chapter.

5.2 Problem Domain & Literature Review

5.2.1 Bright Lesions

Appearance and Distractors

Exudates vary in appearance, conforming to one of three structures: dot exudates, fluffy exudates and circumscribed plaques of exudate. Dot exudates consist of round yellow spots lying superficially or deep in the sensory retina [49]. Fluffy exudates are a paler yellow than dot exudates and tend to lie more superficially in the sensory retina. Plaque exudates vary in size more than the other two

groups and represent a more diffuse accumulation of lipoprotein. Figure 5.1 shows the position of the sensory retina.

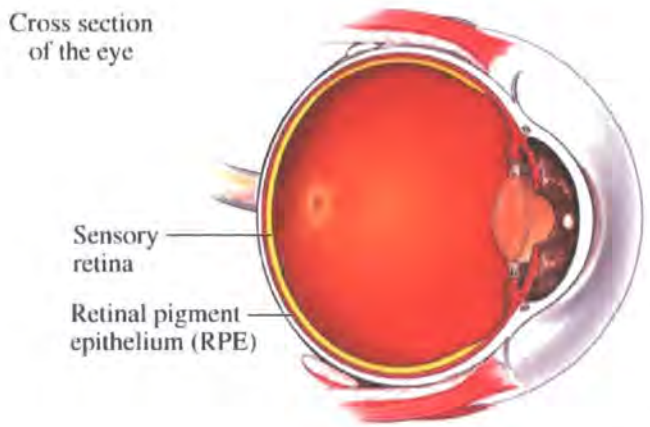


Figure 5.1: Sensory Layer, [taken from http://my.webmd.com/hw/healthguideatoz/hw121946.asp]

In addition to their various appearances, exudates can be arranged in different configurations. Exudates can surround leaking capillaries and microaneurysms in a circular (circinate) pattern or lie scattered, taking on no particular configuration. Exudates are usually reflective and may appear to have a rigid, multifaceted contour, ranging in colour from white to yellow [15]. This wide range in exudate appearance is demonstrated in figure 5.2 where three patches of exudation are shown.

With varying shapes, sizes, patterns and contrast, exudate classification is a demanding problem domain, complicated by lighting variation over the image, natural pigmentation, the intrinsic colour of the lesion, and decreasing colour saturation at lesion boundaries [50]. Reliable identification and classification of exudates is made surprisingly difficult with image distractors such as laser scars and light artefacts. Hard drusen caused by macular degeneration can also add to the complexity, often being mistaken for exudates by naïve approaches due to their similar characteristics.

The similarity between distractors and exudates can been seen in figures 5.2 and 5.3, where drusen, light artefacts and laser burn can be visually compared to the exudate examples.

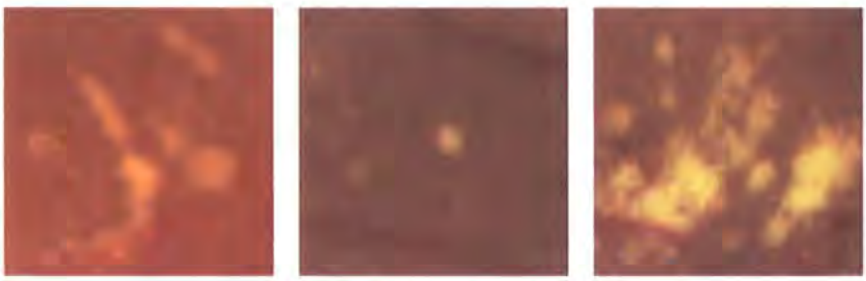


Figure 5.2: Exudates a) Circinate b) Single dot c) Multiple

Figure 5.3: Distractors a) Drusen. b) Light Artefact c) Laser Burn

Segmentation Algorithms

Several authors have presented algorithms for the identification of exudates in fundus images, attaining varied results. Ward *et al.* [51] introduced a semi-automated exudate detection and measurement method, in which an operator selected a threshold value to segment exudates from a shade-corrected retinal background. Ward's method worked adequately well when exudates were the only pathologic feature in the image. However, when faced with other lightly coloured lesions such as cotton wool spots or drusen, the algorithm failed to distinguish between them due to their similar intensity values.

Sinthanayothin *et al.* [8] presented a recursive region-growing algorithm applied to a contrast enhanced image. To reduce the effects of uneven illumination over the fundus, images were pre-processed to enhance local contrast. A prerequisite of local contrast enhancement is a suitable colour model. With the intensity component of the IHS (Intensity Hue Saturation) model being

decoupled from colour, the fundus images were converted from RGB (Red Green Blue) to IHS. Prior to IHS conversion, RGB components were normalized. IHS colour model conversion was achieved using equations 5.1 to 5.3 respectively.

$$I = \frac{1}{3}(R + G + B)) \tag{5.1}$$

$$S = 1 - \frac{3}{R + G + B}[min(R, G, B)] \tag{5.2}$$

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{\frac{1}{2}}}\right\} \tag{5.3}$$

Local contrast enhancement was calculated using sub-windows of 49 × 49 centred on each pixel (i, j) of the ISH's intensity component. The contrast enhancement transformation is defined by equation 5.4 and used a sigmoidal function defined by equation 5.5. The maximum and minimum intensity values of the the intensity image are denoted by f_{max} and f_{min} respectively, with the mean and standard deviation of the intensity within the sub-window W being denoted by $\langle f \rangle_w$ and $\sigma_w(f)$ (equations 5.6 and 5.7) respectively.

$$f(i, j) \rightarrow g(i, j) = 225\frac{[\Psi_w(f) - \Psi_w(f_{min})]}{[\Psi_w(f_{max}) - \Psi_w(f_{min})]} \tag{5.4}$$

$$\Psi_w(f) = \left[1 + exp\left(\frac{\langle f \rangle_w - f}{\sigma_w}\right)\right]^{-1} \tag{5.5}$$

$$< f >_{w(i,j)} (f) = \frac{1}{M^2} \sum_{(k,l)\epsilon W(i,j)} f(k, l) \tag{5.6}$$

$$\sigma_w^2(f) = \frac{1}{M^2} \sum_{(k,l)\epsilon W(i,j)} (f(k, l) - \langle f \rangle_w)^2 \tag{5.7}$$

By using a sigmoidal function, areas with poor initial contrast (small σ value) were greatly enhanced, leaving areas with good initial contrast (large σ value) largely unaltered. To reduce the

noise produced by contrast enhancement, a 2D Gaussian smoothing filter was first applied to the image.

Recursive region growing was applied to the contrast enhanced intensity image with an aim of determining exudate boundaries. Recursive region growing works by dividing the image into regions of similar gray levels. Starting with a pixel at coordinates (x, y) the intensity values of its four neighbouring pixels $(x, y+1)$, $(x, y-1)$, $(x-1, y)$ and $(x+1, y)$ are compared. If the intensity difference between the current pixel and a neighbouring pixel is less or equal to a threshold (in this case 10 intensity levels) it is added to the current region. A region is defined when no more pixels qualify to join the current region. At this point, Sinthanayothin replaces the original intensity values within the region with the median intensity of the area. To split the image into exudate and non-exudate regions, a threshold is set to median intensity of the image background (based on the region with the most pixels). Regions with a median intensity above and below this threshold were classified as exudates and non-exudates respectively. Sinthanayothin stated that the algorithm would not detect faint exudated regions, nor distinguish between other similar coloured lesions.

Goldbaum *et al.* [50] used a quadratic discriminant function to discriminate between similarly coloured lesions (exudates, cotton-wool spots, and drusen). Images were represented by a three dimensional vector (L, Θ, φ), derived by spherical-coordinate transformation (described fully in Wang *et al.* below). Goldbaum observed that exposure noise obscured the radial coordinate (luminance denoted by L) of the colour signal and consequently discarded the luminance component. Classification was accomplished by comparing the two-dimensional chromaticity vector from an unknown object to the two-dimensional chromaticity vector of known objects (training set). The success of this approach was moderate, with Goldbaum concluding that using colour alone was not sufficient to reliably distinguish between similarly coloured lesions.

In a similar approach, Wang *et al.* [52], incorporated image exposure as well as colour information to represent lesions. A discriminant function was used to differentiate between exudates, other lesions and distractors. Lesions were defined by a set of colour features $f_i, f_2, ..., f_k$ forming a k-dimensional space, $\mathbf{F}$. If the ideal features were selected it would possible to find a space $\mathbf{F}$ where objects map to different, non-intersecting clusters.

Wang *et al.* [52] defines a feature space $\mathbf{F}$ to include colour and exposure information only. Although colour fundus images usually consist of three colour planes - red, green and blue, with 256

levels of intensity denoted as (R,G,B) [52], colour can also be represented by spherical coordinates as expressed by:

$$L = (R^2 + G^2 + B^2)^{\frac{1}{2}} \tag{5.8}$$

$$\Theta = Arccos\left(\frac{G}{R}\right) \tag{5.9}$$

$$\varphi = Arccos\left(\frac{B}{L}\right) \tag{5.10}$$

The brightness of the image is denoted by $\mathbf{L}$ and the changes of colour denoted by φ and Θ. The feature space $\mathbf{F}$ is therefore represented by $\mathbf{F}(f_L, f_\Theta, f_\varphi)$. Wang et al, defined two classes - bright yellowish and dark reddish. A training set for each class is obtained by selecting small sub-windows inside exudate and background regions. The means of each exudate and background sub-window are calculated and stored as feature centers for the two class $\mathbf{C}_{lesion}(f_L, f_\Theta, f_\varphi)$ and $\mathbf{C}_{bkgrnd}(f_L, f_\Theta, f_\varphi)$ respectively.

For each pixel in the fundus image the illumination and colour information are extracted $\mathbf{X}(X_L, X_\Theta, X_\varphi)$ and the minimum distance discriminant $D_i(\mathbf{X})$ (equation 5.11) is calculated for each class $D_{lesion}(\mathbf{X})$ and $D_{bkgrnd}(\mathbf{X})$.

$$D_i(\mathbf{X}) = (\mathbf{X} - \mathbf{C}_i)^T (\mathbf{X} - \mathbf{C}_i) \tag{5.11}$$

If $D_{lesion}(\mathbf{X})$ is less than $D_{bkgrnd}(\mathbf{X})$, the pixel $\mathbf{X}$ is classified as lesion; otherwise it is classified as background. Wang's problem domain is less demanding than Goldbaum's, with Wang's emphasis on differentiating yellowish lesions with darker objects as opposed to distinguishing between similar coloured lesions. However, Wang presents promising results within the limited image dataset. This technique is therefore compared against other techniques latter in this chapter.

Osareh *et al.* [53] introduced a fuzzy C-Means clustering algorithm based on the work of Young Won Lim *et al.* [54] to segment a colour retinal image into homogenous regions; classifying regions into exudates or non exudates using a neural network. To compensate for the wide variation of colour in the fundus, the images were pre-processed. Images were converted from RGB to IHS,

normalised and finally locally contrast enhanced (described above in Sinthanayothin *et al.*)

Fuzzy C-Means clustering allows pixels to be classified into multiply classes with varying degrees of membership. Young Won Lim's algorithm consisted of two stages - coarse and fine segmentation. The coarse segmentation divides the image into a number of regions determined by thresholds automatically generated by scale-space filtering Hue Saturation and Intensity (HSI) histograms. The number of significant peaks in the histogram determine the number of regions. Pixels not segmented by the coarse segmentation phase are processed in the fine segmentation phase by determining fuzzy partitioning. This is done by minimizing the squared error loss function 5.12.

$$L = \sum_{j=1}^{m} \sum_{i=1}^{n} [\mu_j(x_i)]^b \|x_i - c_j\|^2 \tag{5.12}$$

where x is a data vector $(i = 1, ..., n)$,n is the number of pixels, c_j denotes the centre of fuzzy clusters $(j = 1, ..., m)$ where m is the number of clusters determined in the coarse segmentation stage. The fuzzy membership of x_i to cluster j is denoted by $u_j(x_i)$, and the overlap between the fuzzy cluster regions is controlled by a weighting exponent b. The similarity between each region's cluster centre and pixel attribute vectors are measured by equation 5.12. Osarch *et al.* performs differentiation and applies the constraint $\sum_{j=1}^{m} \mu_i(x_i) = 1$ resulting in functions 5.14 to 5.15.

$$\mu_j(x_i) = \left[\sum_{k=1}^{k=m} \left(\frac{d_{ji}}{d_{ki}} \right)^{\frac{1}{b-1}} \right]^{-1} \tag{5.13}$$

$$d_{ji} = \|x_i - c_j\| \tag{5.14}$$

$$c_j = \frac{\sum_i [\mu_i(x_i)]^b x_i}{\sum_i [\mu_i(x_i)]^b} \tag{5.15}$$

The Euclidean distance between the cluster centre and the data vector is denoted by d_{ji}. Osarch initialised the weighting exponent b to 2 and minimised the FCM function 5.14 until the change in the norm was less than 0.5 and three different clusters were distinguished - dark vessels and haemorrhages, background pigmentation and exudates. Osarch et al. states that the segmentation by FCM is a conservative process finding all but the faintest (ambiguous) exudate regions. False positive non-exudate segmented regions were also found by the algorithm caused by cluster

overlapping, noise, and uneven colour distribution. After segmentation, exudate and non-exudate regions were classified as being exudate and non-exudate by extracting a number of region features (size, colour, average intensity, edge sharpness and standard deviation of intensity), and feeding the information into a multilayer neural network with conjugate gradient decent as a learning method. Osarch used 42 images containing 4037 objects for both training and testing. The training set consisted of 3000 objects, 1205 of them labelled as exudates. Of the remaining 1037 objects 417 were labelled as exudates. From this limited dataset Osareh reported a sensitivity of 92% and specificity of 82%.

5.2.2 Dark Lesions

Appearance and Distractors

Microaneurysms and haemorrhages often coexist within fundal images - microaneurysms represent saccular bulges in the walls of the retinal capillaries [55] and haemorrhages are thought to occur from ruptured bulges. Found in any part of the retina, microaneurysms have a rigid structure appearing as small, round, red dots with a limited radius range. Haemorrhages, however, vary more in appearance. At their smallest, haemorrhages can be difficult to distinguish from microaneurysms. Haemorrhages can have irregular boundaries being either "flame shaped" or "dot and blot" in appearance. The size of haemorrhages can also differ, ranging from areas the size of microaneurysms to areas larger than the optic nerve head. Figure 5.4 shows fundus areas with microaneurysms, and the varying appearances of haemorrhages.

Figure 5.4: Red Lesions a) Haemorrhage b) Microaneurysm c) Mixed

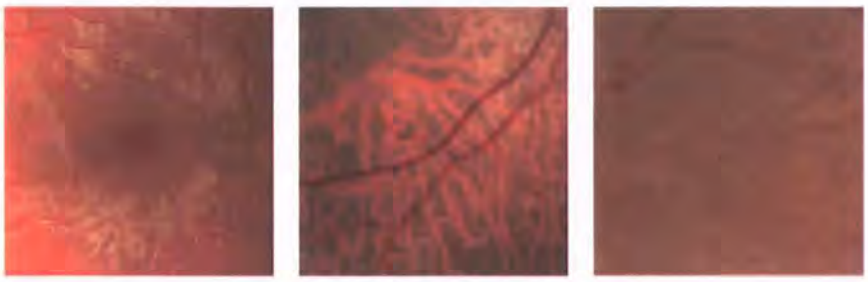


Figure 5.5: Distractors a) Light Artefact b) Choroidal vessels c) Vessel Breaks

With microaneurysms and haemorrhages being the same colour as blood vessels, vascular distractors make the segmentation of red lesions a challenging problem domain. When veins change angle within the retina or arteries cross, fine veins can appear broken with dark followed by lighter segments. Combined with vascular light reflex (bright line running through blood vessels), segments of the vascular network are a major distractor appearing as a number of microaneurysms or small haemorrhages (see figure 5.5 c).

Light artefacts and varying natural pigmentation can also play their part in complicating segmentation. Figure 5.5 a. shows a large light artefact around the fovea, resulting in pockets of normal pigmentation surrounded by lighter areas. This can cause problems in selecting candidate dark lesions as the normal pigmentation appears dark in relation to its local surrounding area. The same effect can be seen in Figure 5.5 b., where the retinal pigment has visible choroidal vessels ("tiger stripes").

Segmentation Algorithms

A number of authors have developed techniques to segment and classify microaneurysms and haemorrhages from fundus images. The majority of algorithms use similar pre-processing and region growing methods.

Spencer *et al.* [56], pre-processed high quality digitised fluorescein angiogram negatives (capturing frames at 1024×1024 resolution) by shade correction and used a mixture of morphological opening, matched filtering and thresholding to identify candidate microaneurysms. Microaneurysm seeds were region grown and finally classified using a rule set based on features extracted from the

grown region.

Spencer *et al.* [56] used shade correction to help separate the microaneurysms from the retinal background. This was achieved by approximating the retinal background and removing it from the original. The approximated background image was obtained by applying a median filter whose size was said to be larger than a microaneurysm. Microaneurysms were then segmented from the vasculature by morphologically opening the image with a linear structuring element set to different orientations. To ensure a high degree of discrimination between microaneurysms (circular) and vessels (linear), the length of linear structuring element was set greater than the largest microaneurysm (11-pixels). To obtain good alignment with vessel segments but also maintain computation to a minimum, the structuring element was set at eight different rotational orientations - 0, 30, 45, 60, 90, 120, 135 and 160 degrees. The maximum pixel value at each spatial location from the eight opened images were combined to produce a vascular image with no circular features. This resultant image was then subtracted from the shade corrected image to produce an image with in theory only circular features. Remaining unwanted features were removed from the vascular segmented image by applying 11×11 Gaussian filter ($\sigma = 1$). Images were thresholded by a fixed number of pixels above the modal grey-level to establish potential lesions seed points. At each seed location, recursive region growing was used to determine the boundary of the potential lesions. Features relating to boundary shape and size were extracted from region grown areas; microaneurysm classification was achieved by assessing candidate features with series of rules contained within a rule set. Spencer *et al.* [56] achieved a sensitivity of 82% at the cost of detecting over 100 false positives over four images.

Although fluorescein angiography is an extremely effective method of microaneurysm detection, and studies have shown automated techniques in this area [Spencer], angiography is an invasive procedure and not part of routine diabetic retinopathy screening. However, techniques used in this area are transferable and have been adopted by other authors using standard fundus photographs.

Oien *et al.* [1], introduces an algorithm to segment microaneurysms from 22 high resolution digitally scanned fundal photographs (1536×1024). During the photographic or scanning process, unwanted particles such as dust or dirt may adhere to the image, although typically smaller and of lower intensity than microaneurysms, false positive classifications are possible. To remove any small dark spots that might be extraneous particles while leaving the majority of the image unaltered,

a median filter with quadratic mask (width twice the size of expected particles) was applied. Candidate extraneous areas were determined by differences between the original and filtered image. These areas were then confirmed by thresholding with a low intensity (30 pixels). Pixels within confirmed extraneous areas were replaced by estimated pixel values from a median filter (11×11 pixels).

In a similar approach to Spencer *et al.* [56], the background's nonuniform intensity (caused by uneven intensity of illumination), was reduced using shade correction; applying a large median filter (31×31 pixels) followed by the subtraction of the filtered image from the original. In order to isolate candidate microaneurysms, percentwise thresholding was applied to the pre-processed difference image. In this technique, a percentage of pixels with the lowest intensity values are converted to black with the remaining pixels converted to white. Oien selected percentage values to create two binary images - 10% for microaneurysms detection and 15% for vessel detection.

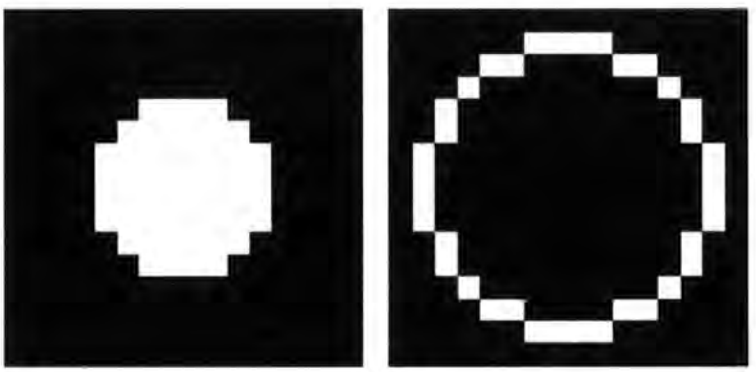


Figure 5.6: Hit-and-miss structuring elements a) Object. b) Background.

Circular or near circular spots in the 10% binary image were classified as microaneurysms using a *Hit-and-miss* algorithm. In this method two structuring elements representing the object and the background are eroded from the image (see figure 5.6). The intersection of these two erosions provides the location of objects with a geometry and size as defined by the two structuring elements. Three pairs of structuring elements were used to detect small (3-7), medium (5-11) and large (9-17) microaneurysm. In the final stage of the algorithm, blood vessels were morphologically eroded from the 15% binary image and subtracted from the microaneurysm image, ensuring no false hits

occurred on the vascular network. This morphological erosion involved a linear structuring element rotated to 16 spatial angles. The detected areas not included in the segmented vascular network were deemed to be microaneurysms. Individual microaneurysm detection was reported at 72% with a false positive rate of 60%. Imperfection in the vessel segmentation was said to account for the majority of the false positives. Therefore a more sophisticated vascular segmentation algorithm could dramatically improve performance of this approach.

Developing upon the microaneurysm segmentation work of Spencer *et al.* [56], Hipwell's *et al.* shade-corrected images prior to the removal of blood vessels and haemorrhages - achieved by excluding all structures greater than a particular linear extent. Measurements including intensity, shape, perimeter length, aspect ratio, and circularity were extracted from candidate areas resembling microaneurysms. Candidates were classified as microaneurysms if measurement features passed a series of tests. Hipwell et al reports a sensitivity 81% and specificity of 93%. An accurate comparison cannot be made with Hipwell's results due to the way the sensitivity and specificity was calculated. Hipwell evaluated on a per patient basis as opposed to per image. Patients were graded after examining four images, two images per eye.

Sinthanayothin *et al.* pre-processed images with a high pass filter to enhance the contrast between lesions and background. A recursive region-growing algorithm identified candidate microaneurysms and haemorrhages. Due to intensity similarities, blood vessel segments were often classified as lesions. To overcome this problem a neural network was used to segment the vascular network from the image. Sinthanayothin states the algorithm failed to find haemorrhages adjacent to blood vessels and falsely identified small retinal vessels as red lesions.

5.2.3 Classification

Classification can be defined by two distinct approaches – given a set of observations, classes are established to cluster the data; or given the number of classes, a rule is established to classify a new observation into one of the existing classes. The former approach is known as unsupervised learning and the latter as supervised learning [57]. Lesion classification techniques have largely focused on supervised learning as the number of lesion types or classes are known. Supervised learning can be achieved using a number of approaches including: statistical, machine learning and neural network.

1. **Statistical approaches** Statistical approaches can be generalised as having an underlying probabilistic model, in which the probability of a new observation belonging to each class is used as the classifier [58].

2. **Machine learning** The machine learning umbrella incorporates a wide variety of techniques all focused on mimicking human reasoning. From this umbrella, two distinct machine learning factions have emerged: rule-based, in which classification results from a sequence of logical steps and data analysis, where rule-structured classifiers are derived from pre-classified examples [58]. A useful prerequisite of machine learning techniques is that the learning concepts should be in a symbolic form that is meaningful to humans.

3. **Neural Networks** Neural networks combine features from both statistical and machine learning models to produce a probabilistic model that aims to mimic human reasoning. The idea behind neural networks is to capture the important features of a biological neural system in a model in order for it to exhibit similar behaviours. Neural networks are comprised of many interconnected units (perceptrons) each of which performs a weighted sum of its inputs. Layers of interconnected perceptrons can represent a complex set of interdependencies allowing nonlinear problems to be modelled. In order for a perception neural network to be useful, the network needs to be able to learn. This is achieved by strengthening or weakening connection weights depending upon the validity of its output prediction. A drawback of neural networks is that the internal neural model is not transparent, therefore the learnt concepts of the model are unknown. This can cause a problem if the training data is not truly representative of the actual data – as abstract data features may yield more classification influence than it otherwise would.

Previous authors have experimented with all three types of supervised learning in comparative studies. Frame *et al.* [55] compares three microaneurysm classifying techniques using Spencer's *et al.* [56] pre-processing and region growing algorithim. A rule based system, linear discriminant analysis and a learning vector quantisation artificial neural network were all compared. The three classifiers were trained using a set of 1659 candidate lesions, gathered from 68 retinal fluorescein angiographic images, with 400 labelled as microaneurysms. An independent test set consisting of 1067 candidate lesions, (containing 297 classified microaneurysms) from 20 images, validated

results from the three classifiers.

1. **Rule Based Analysis:** The rule based classifier distinguished between lesion and non-lesion by satisfying a number of logical rules derived directly from the training data. Pairs of features were compared from feature vector scatter graphs to find two feature functions that gave good discrimination between lesion and non-lesion classes. To avoid conflict-resolution strategies, all rules were mutually exclusive. Generating a rule set can be extremely time consuming, Frame *et al.* [55] dedicated one man month to discriminating between microaneurysm and spurious objects.

2. **Linear Discriminant Analysis:** Linear discriminant analysis is a statistical classifier that maximises separation between classes. This algorithm seeks directions for efficient discrimination between classes, placing a hyperplane at an optimum location in feature space to achieve maximum separation. [59].

3. **Learning Vector Quantisation Artificial Neural Network:** Learning vector quantisation is a competitive learning two layer neural network encompassing an input and output layer. The network includes a set of parameter vectors (codebook vectors) whose coordinates represent the weights of the connections between an input and the output layers. Each codebook vector is associated with an output class. Each class may have a number of codebook vectors and contain the same number of elements as feature input to the network (input vector). The algorithm works by comparing the input vector with each of the codebook vectors. The vector with the closest Euclidean distance is the winner and the associated class taken as the algorithm's output. The training set is used to assign codebook vectors to the correct class by moving winning vectors closer to the input vectors.

Frame concluded that the rule based system required further development but gave higher performance than the other two methods.

Ege *et al.* [60] presented work on three different statistical lesion classifiers, comparing a K Nearest Neighbour (KNN), Mahalanobis, and Bayes classifier. For each classifier type, two classifiers were constructed - one for bright lesions and one for dark lesions. The data was also split equally into two sets; a learning set and a test set.

1. **K Nearest Neighbour:** In order to differentiate between objects, the K Nearest Neighbour classifier is presented examples of objects representative of each class. Features F representing these objects are extracted and form a training set. Each feature within the training set is assigned a class label, C=c1,...,cn. The Euclidean distance is assigned between all feature points in the training set, forming a distance matrix between all possible pairings.

   To determine the class of an unseen object, the distance matrix is analyzed to find the k-closest neighbors to the object's feature points. The k-closest feature points are then analyzed to determine the class label majority. The most common class is assigned to the object being analyzed [61]. The size of K is equal to $\sqrt{n}$ where n is the total number of samples in the training set [60].

2. **Bayes Classifier:** Based on the Bayes decision rule, the Bayes classifier can be used to minimise the expected total number of errors [60]. The Bayes theorem is expressed by

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(x)} \tag{5.16}$$

   where

$$p(\mathbf{x}) = \sum_{i=1}^{n} p(\mathbf{x}|\omega_i)P(\omega_i) \tag{5.17}$$

   For each candidate lesion, a feature set consisting of nine variables was extracted and denoted by $\mathbf{x}$. $P(\omega_i|\mathbf{x})$ is the posterior probability that the feature set $\mathbf{x}$ belongs to class i. $p(\mathbf{x}|\omega_i)$ is the state-conditional probability density function for $\mathbf{x}$. The prior probability of a lesion class i occurring in an entire retinal imageset is denoted by $P(\omega_i)$ and calculated by equation 5.18.

$$P_{w_i} = \frac{f_i}{n} \tag{5.18}$$

   where f_i and n represent the observations of lesion class i and the total number of observations respectively. In Ege's work, if the candidate lesion was bright, it was tested against the exudates or cotton wool spots class, where the value of class i was 1 or 2 respectively, and

the total number of classes n was 2.

To classify a candidate lesion, the posterior probability of both classes requires calculation. A decision is made in favour of ω_1 (exudate) if the expression in equation 5.19 is fulfilled.

$$P(\omega_1|x) > P(\omega_2|x) \tag{5.19}$$

Due to the multivariate normally distributed data (9 variable feature set), $p(x|\omega_i)$ is calculated using the state-conditional probability density function as expressed in equation 5.20

$$P(\omega_1|x) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\sum_i|}} exp\left(\frac{1}{2}(x-\mu_i)^T \sum_i^{-1}(x-\mu_i)\right) \tag{5.20}$$

where d is the number of features, μ_i represents the mean of class i and $\sum_i$ denotes the covariance matrix of class i.

3. **Mahalanobis Classifier** The Mahalanobis classifier is a similar idea to the nearest neighbour classifier with properties of the Bayes classifier. The distance metric of this approach is the Mahalanobis distance and the parameters mean and covariance are the same as the Bayes classifier. The Mahalanobis classifier does not use a *priori* probability of classes but does assume multivariate normally distributed data. The Mahalanobis distance is expressed in equation 5.21

$$r_i^2 = (x-\mu_i)^T \sum_i^{-1}(x-\mu_i) \tag{5.21}$$

where x is the feature set from the candidate lesion being tested, μ_i the mean of class i and $\sum_i$ is the covariance matrix of class i. The tested candidate lesion is assigned to the class returning the lowest distance. Ege states that the Mahalanobis classifier gave the most balanced results in their 134 image test set, but concludes that the difference between the Bayes and Mahalanobis classifier may be due to the sparse training data.

Gardner *et al.* [62] used a multilayer perceptron neural network to classify 20×20 sub-windows from red-free fundus images as exudates, haemorrhages vessels, or normal. After median smoothing,

red free values were fed directly into a large neural network comprising of 400 inputs and 80 hidden units. With each pixel corresponding to a neural input, the approach is computationally demanding. Gardner reports sensitivity and specificity of 99% and 69% respectively for patient referral with either exudates or haemorrhages. Gardner did not elaborate on how exudates and haemorrhages were encoded and integrated to come to this outcome.

Hunter *et al.* [63] [64], presented an algorithm to distinguish exudates and drusen from tiled fundal images using a multilayer perceptron neural network, and introduced a novel approach to feature selection using sensitivity analysis. Red-free fundal images were divided into 16×16 pixel tiles. Fast Fourier transforms and Prewitt edge-detection filters were applied to the fundus images, resulting in three images transforms (red-free, fourier and prewitt). For each image transform, the following features were extracted: summary statistics (Mean, Standard Deviation, Skew and Kurtosis), principal components analysis, 16-bin histogram, and principal components of histogram. With over 156 features extracted from each tile, selecting the right combination of features is critical in reducing the complexity of the neural network, increasing performance and increasing execution speed.

Hunter's algorithm is an hierarchical approach, involving feature selection stages that identify significant features. In each stage, a neural network is trained, and analysed showing the contribution of each feature. Using this technique composite feature subsets can be selected among hundreds of features to create a robust network with optimum performance. Hunter created 4 networks (summary statistic , principal components analysis, 16-bin histogram, and principal component of histogram) for each of the three image transforms. Sensitivity analysis was applied to each network respectively, creating a composite network for each image transform. Each composite was further analysed and combined to create an overall composite, and re-analysed to create a final network of 11 inputs. To train the networks a data set containing 95 drusen tiles and 116 exudate tiles were split into a training set (44 drusen, 56 exudate) and a test set (51 drusen, 60 exudate). Hunter reported performance of 91% correct classification from the final network using the test set. This compared to 60% for a network trained directly from pixel values.

5.2.4 Summary

The prevention of sight threatening retinopathy (maculopathy) requires the early detection of various clinical features within the retina. These features include exudates (bright lesions), microaneurysms and haemorrhages (dark lesions). The detection of both bright and dark lesions is complicated by image distractors that have similar appearances and intensity values to the lesions.

Bright and dark lesions have a combination of core and unique image distractors that require separate segmentation algorithms. With bright lesions, fuzzy C-Means clustering, a colour discriminant function, and a recursive region growing algorithm have all been used to segment exudates from the retinal background. These approaches achieve good sensitivity by correctly identifying exudates but have a poor specificity by misclassifying similar coloured distractors as exudates. With dark lesion segmentation, images were largely shade corrected prior to the removal of blood vessels. Microaneurysm and haemorrhage detection was achieved by recursive region growing and thresholding techniques. Both approaches suffered from false positive results on unsegmented segments of the vascular network.

Statistical, machine learning and neural network techniques have been employed for lesion classification. Neural networks combine features from both machine learning and statistical models. Selecting suitable lesion descriptors can be a difficult and time consuming process. Sensitivity analysis allows a liberal feature selection, by identifying the descriptors that contribute to the neural network.

5.3 Method

The accurate location and classification of diabetic lesions is the primary goal of this research, as early detection of exudates, microaneurysms and haemorrhages within the macula, can prevent sight threatening retinopathy if treated soon after diagnosis. The location and segmentation of structural elements of the retina including the optic nerve head, fovea, and blood vessel network are all prerequisites for this phase of retinal analysis. Optic disc and fovea segmentation are required to locate the centre and size of the macula region (where lesion presence is the most significant) and blood vessel network segmentation reduces red lesion image distractors.

In this research, lesion identification has been divided into three stages: location, segmentation

and classification. The main contribution to novelty exists in the segmentation phase, with some contribution in the classification phase.

This section begins with a brief description of the lesion location algorithm used (but not developed) in this research and is presented for completeness. This is followed by a detailed description of the new region growing algorithm that is used to segment lesions from the retinal background. An explanation of the neural classification phase of the algorithm concludes this section.

5.3.1 Location

Segmenting a lesion from the retinal background is a computationally demanding process. To reduce computational expense, only the most likely candidates should be segmented. This is achieved by first locating all possible lesion candidates. Point (a.k.a. peak) detection is used to identify isolated points where the intensity value of the isolated point is quite different from its neighbouring pixels. A small 5×5 sub-window is morphologically run over the fundus image, applying a maximum (minimum for dark lesions) filter within each sub-window, producing peak/trough points.

As large numbers of peak/trough points exist per image and the majority are within image distractors such as lighting variation, natural pigmentation, light artefacts, and broken capillaries, a simple test is used to sort these points into an order of lesionness.

By exploiting lesion characteristics (as detailed description in section 5.2.1 and 5.2.2), shape, size and contrast details can be used to grade these peak points. If it is assumed that a lesion is a roundish object with high contrast and a diameter between 2 and 30 pixels, a series of seven concentric circles, can measure the contrast at different radii between peak and retinal background. The highest contrast between concentric circles and peak is taken as a seed's lesionness value. As an image contains many distractors some of which will be circular with high contrast, many points require further investigation to ensure actual lesions are among those analysed. To this end, the top 100 sorted peak points are selected for region growing, feature extraction and classification.

5.3.2 Segmentation - Region Growing

Region growing is a technique that groups pixels together to form larger regions. Pixel aggregation is one the simplest of these approaches in which regions are grown from a set of "seed" points by appending neighbouring pixels that share similar properties to the seed. These properties may include intensity, texture or colour [12]. Pixels are repeatedly appended, forming increasing regions until a stopping criteria is satisfied.

The proposed method uses traditional intensity pixel aggregation for region growth while iterately calculating the contrast between the average boundary and average grown region gradient. The algorithm continues to grow exceeding the objects natural boundary, halting only when the grown region loses its compactness. The objects boundary is then locating by determining the point of growth at which the objects contrast gradient is most significant.

Growing

The region contrast gradient growth model in its most basic form consists of two iterately growing regions - core and boundary. The core is initially grown from a "seed" point and continually grows by appending the brightest neighbouring (boundary) pixels on each iteration. Unlike other region growing algorithms, the inclusion of pixels does not depend on satisfying a threshold. As the intensity difference between the background retina and lesion seed is so variable, a fixed threshold is unreliable. Instead, the highest intensity pixel is iteratively grown regardless of the seed intensity.

To illustrate this procedure, consider figure 5.7a. where the numbers inside the cells represent intensity values from the intensity component of an HSI image R. Let point S_1 with coordinates $(4,4)$ represent a seed point and E denote a 3×3 structuring element. Let intensity image R be partitioned into n subregions where n denotes the number of seed points $R_1, R_2, ...R_n$. In this example $n = 1$ and region R_1 is initialised to seed point S_1. Region R_1's neighbouring pixels B_1 (as highlighted in red in 5.7a.) are determined by dilating R_1 with E as expressed in equation 5.22, followed by subtracting R_1 from $R_1 \oplus E$ as depicted in equation 5.23.

$$E = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$$

$$R_i \oplus E \underline{\triangle} \{x : E_x \cap R_i \neq \phi\} \tag{5.22}$$

$$B_i = R_i \oplus E - R_i \tag{5.23}$$

where x is the set all pixels of R_i and E_x is a translation of E to origin x. Within the boundary B_1, the pixel with highest intensity value $(5, 3)$ is assigned to R_1. The boundary of the appended region R_1 is determined by dilation as described above, with resultant region and boundary shown in figure 5.7b; with the growth of the third pixel and resultant dilated boundary shown in figure 5.7c. The inclusion of the brightest pixel from boundary B_1 into the region R_1 is repeated until a stopping criteria (as described below) is met.

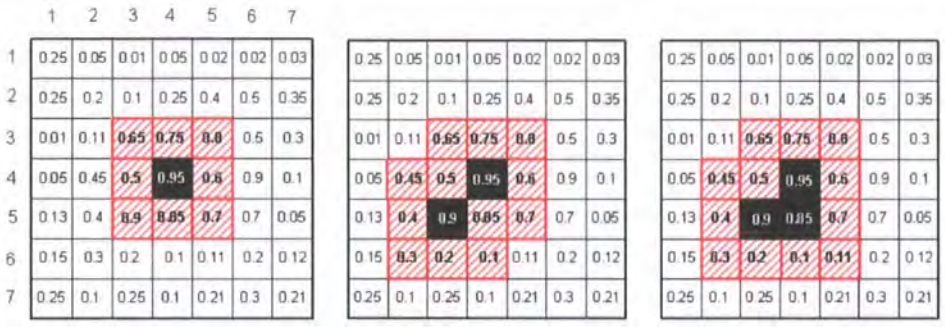


Figure 5.7: Example of region growing a) iteration 1 b) iteration 2 c) iteration 3

Gradient Measure

The boundary of a lesion can be characterised by the point of strongest intensity contrast between itself and the background retina. By determining the gradient of image R, this maximum rate of change can be exploited. Prior to gradient conversion, the intensity image is gaussian-smoothed to reduce overall image noise and to bridge any small gaps within the lesion boundaries. The image is also normalised to increase the images dynamic range. Segmentation is therefore achieved using a normalized negative gaussian-smoothed gradient magnitude image, $\Gamma = -\|\nabla \mathbf{I}\| / \max(\|\nabla \mathbf{I}\|)$.

The magnitude of the gradient image can detect the edge amplitude at which pixels change their gray-level suddenly.

The gradient of pixels situated on a lesion edge is high. The gradient contrast between boundary and core is therefore strongest when core pixels are positioned on this edge. To determine an optimum boundary, gradient measures are taken at the inclusion of each new pixel, followed by a measure of gradient contrast between boundary and core. At region iteration j, of region i, the mean gradient pixel values of image Γ masked with region R_i and boundary B_i are determined by equation 5.24 and 5.25 respectively. The gradient contrast between the boundary and core is expressed by equation 5.26.

$$\phi_{i,j} = \sum \frac{R_i \cdot \Gamma}{\tau} \cdot \omega_i \tag{5.24}$$

$$\beta_{i,j} = \sum \frac{B_i \cdot \Gamma}{\kappa} \cdot \omega_i \tag{5.25}$$

$$\zeta_{i,j} = \beta_{i,j} - \phi_{i,j} \tag{5.26}$$

where κ, τ denote the pixel count of B_i and R_i at iteration j respectively. The contour gradient around seed points with especially high intensity values can occasionally have gradients greater than that of the true lesion edge. To reduce the risk of inaccurate boundaries around seed points, a diminishing weight ω_j is multiplied to the average gradient, reducing the gradient impact during the first ten iterations (see equation 5.27).

$$\omega_j = 1.0 + 0.4^j \tag{5.27}$$

Once the stopping criteria is fulfilled and the region growth is terminated, the optimum boundary is sought. The optimum boundary is identified by the growth iteration with the maximum contrast between boundary and core, as expressed by equation 5.28.

$$\psi_i = \arg_j \max(\zeta_{i,j}) \tag{5.28}$$

The region growth procedure is then repeated, stopping after ψ_i iterations. At this point, region features are extracted.

Stopping Criteria

A significant problem of region growing algorithms is the formulation of a suitable stopping criteria. A region should stop growing when no more pixels satisfy the condition for the inclusion to that region. If the intensity difference between a lesion and background is marginal, a stopping criteria based on an intensity percentage or threshold from a seed's intensity can yield problems. More advanced stopping conditions may include additional rules specifying the maximum size and or deformation from a template shape.

A number of problems exist with using such stopping rules when applied to lesion segmentation; firstly, a segmented boundary that has been terminated due to excessive size or shape deformation is unlikely to be an optimum lesion boundary. Secondly, lesions vary significantly in both shape and size, so defining a termination rule based on either of these features would likely produce varied and unsatisfactory results.

Compactness, a dimensionless quantity that is insensitive to scale and orientation changes is presented as the algorithm's stopping criteria. The compactness of a region is defined as $perimeter^2/area$, where the area of a region is defined by the number of pixels contained within its boundary and the perimeter of a region is defined by the length of its boundary [12].

The compactness of the internal region is determined at each iteration. Retinal background intensity varies greatly, consequently after the the internal region grows to the full extent of the lesion contour, the *ad hoc* inclusion of additional retinal pixels causes the compactness to deteriorate. If the compactness metric exceeds a threshold of 30, region growing is terminated.

5.3.3 Feature Extraction

Having determined the boundary of the lesion candidate, it is important to extract as much information as possible from the region of interest; it is this information that is be utilised to classify the potential lesion.

A region can be represented by a number of characteristics such as shape features (area, perimeter, orientation and concavities) and internal reflectivity (colour and texture). With a large number

of descriptors available, determining the importance of particular features can be a difficult and time consuming process.

Sensitivity analysis, as described in section 5.2.1, can simplify this process by expressing a significance rating for each descriptor. A liberal set of features, including closely related or even overlapping descriptors can be extracted (as feature selection reduces the feature set before use by the classifier). The extracted features include a number of contrast, colour, shape and texture descriptors. Although certain features represent dark lesions more than bright lesions and *vice versa*, the same features were extracted for both, allowing sensitively analysis to identity the significant contributors.

Colour

Colour is a powerful descriptor, often simplifying object identification. Colour is especially important in distinguishing between exudates and similar shaped distractors such as drusen. As the colour differences between exudates and drusen are often so marginal, three different colour models are used to extract colour features.

1. **RGB Model** Images captured from the Canon CR6 45MNf fundus camera are represented in the primary spectral components of red, green and blue. Although the green component contains the majority of visible contrast between lesions and background retina, features from all colour planes are extracted.

2. **CIE Model**

   All colours in this model are defined by hue and chroma, not by brightness. When the luminance is removed from colour, the isoilluminant surface describes the chromaticity part of colour. Both Goldbaum [50] and Wang [52] have exclusively used this model to identify exudates; however, in this research it is used as one of three colour models.

   To convert from RGB to the CIE model the following equations 5.29 to 5.31 are used.

$$\ell = (G^2 + R^2 + B^2)^{\frac{1}{2}} \tag{5.29}$$

$$\theta = \arctan\left(\frac{G}{B}\right) \tag{5.30}$$

$$\phi = \arccos\left(\frac{B}{\ell}\right) \tag{5.31}$$

3. **HSI Model**

The HSI model uses three quantities hue, saturation and intensity to specify colour. Hue is a colour attribute that describes a pure colour (pure yellow, orange, or red) and as saturation is a measure of the degree to which a pure colour is diluted by light. As with the CIE model, the intensity component or luminance is decoupled from the colour. Although there is a feature overlap between CIE and HSI colour models, the quantities are calculated in different ways. To convert from RGB to HSI the three RGB components should be in the range of $[0, 1]$. The intensity, saturation and hue are defined by equations 5.32 to 5.34 respectively.

$$I = \frac{1}{3}(R + G + B) \tag{5.32}$$

$$S = 1 - \frac{3}{R + G + B}[min(R, G, B)] \tag{5.33}$$

$$H = cos^{-1}\left(\frac{\frac{1}{2}[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{\frac{1}{2}}}\right) \tag{5.34}$$

Characteristics of a Frequency Distribution

To identify colour pixel dispersal within a segmented region, the frequency distribution is analysed in each plane of the three colour models. The tendency of the colour regions is examined to determine if pixels are centrally or more uniformly distributed. To measure the tendencies of the frequency distribution statistical moments are calculated. The k^{th} moment of a frequency distribution about any point x_0 is defined by equation 5.35.

$$m_k = \frac{\sum_{i=1}^{n}(x_i - x_0)^k}{n} \tag{5.35}$$

Measurements of central tendency, general variability, symmetry of variation and shape are calculated by the mean, standard deviation, skew and kurtosis respectively. The arithmetic mean is a commonly used measure of central tendency and is the first moment of the distribution about zero (equation 5.36).

$$\mu = \frac{\sum_{i=1}^{n}(x_i)}{n} \tag{5.36}$$

To measure the general variability of a frequency distribution, the standard deviation is calculated. The variance of a distribution is the second moment of a distribution about the arithmetic mean (equation 5.37).

$$\sigma = \sqrt{var(x)} \tag{5.37}$$

$$var(x) = \frac{\sum_{i=1}^{n}(x-\mu)^2}{n} \tag{5.38}$$

The third moment around the mean is taken as a measure of absolute skewness of the distribution, and represents the symmetry around the point of central tendency.

$$\gamma_3 = \frac{m_3}{m_2^{\frac{3}{2}}} \tag{5.39}$$

The shape of the distribution is measured by calculating its kurtosis. If the distribution has a sharp peak, thin shoulders, and fat tails it is said to have a high kurtosis. If the distribution has a flat peak, fat shoulders and thin tails then it has a low kurtosis (equation 5.40).

$$\gamma_4 = \frac{m_4}{m_2^2} - 3 \tag{5.40}$$

In addition to colour frequency analysis, the statistical features were extracted from Fourier transformed and Prewitt edge detected regions.

Principal Component Analysis

Principal component analysis (PCA) is a useful statistical technique to identify patterns within data (image regions). PCA represents the data by aligning a series of axe' in the direction of data variance i.e. the first component is aligned to the largest percentage of the total variance and each subsequent axis contains the maximum variance for any axes orthogonal to the previous component. As the importance of information from each subsequent axis diminishes, only the first 10 principal components are used as image features; these are extracted from the green component of the RGB colour model, Fourier and Prewitt transforms.

To calculate the PCA of an image, the mean of each data dimension is first subtracted from the values in that dimension, producing a data set whose mean is zero. The disparity of each dimension from the mean with respect to each other is determined by calculating the covariance of each dimension pair, thus producing a covariance matrix. The eigenvectors of the covariance matrix extracts lines that characterise the data. Ordering the eigenvectors by eigenvalue, highest to lowest gives the components in order of significance as the eigenvector with the highest eigenvalue is the principal component.

Regional Shape Descriptors

Regional features provide discriminating information between similar shaped regions and background. Regional features extracted include: area, perimeter, major and minor axis lengths, compactness, convex area, eccentricity, solidity, and equivalent diameter.

1. **Area** number of pixels contained within boundary
2. **Perimeter** pixel length of the boundary
3. **Compactness** $perimeter^2/area$ dimensionless value that is insensitive to scale and orientation changes.
4. **Major Axis** length of straight line joining two furthest points.
5. **Minor Axis** length of straight line joining two furthest points perpendicular to major axis
6. **Eccentricity** ratio between major and minor axis

7. **Convex area** area of the smallest convex hull encapsulating the boundary.

8. **Solidity** $Area/ConvexArea$ proportion of pixels in both convex hull and region.

9. **Equivalent diameter** $\sqrt{(4 * Area/\pi)}$ diameter of a circle with the same area as the region.

Contrast

Contrast can be defined as the relative intensity difference between adjacent regions of an image. In this research measures of contrast are extracted from the green component of the RGB colour model. The contrast feature measures the intensity difference between the mean boundary pixels and the mean region pixels of the segmented area; as expressed in equation 5.41.

$$c = |\overline{r} - \overline{b}| \tag{5.41}$$

$$\overline{r} = \frac{1}{n} \cdot \sum_{(k,l)\epsilon\alpha(x,y)} I(k,l) \tag{5.42}$$

$$\overline{b} = \frac{1}{m} \cdot \sum_{(k,l)\epsilon\kappa(x,y)} I(k,l)$$

(5.43)

where α, and κ represent the pixels and n and m represent the pixel count of the region boundary and boundary respectively.

Texture

Texture content quantification is an important region descriptor. Texture definition measures such properties as coarseness, smoothness and regularity. From the three principal approaches of textural analysis, statistical, structural and spectral, statistical analysis yields characterisations of texture that are vitally important in discriminating between different types of lesions, such as grain content, smoothness and coarseness. The relative position of the pixels with respect to each other can be used to extract such information. To obtain the complete picture regarding pixel positions

four co-occurrence matrixes are needed, one for each directional join to the next pixel i.e. right, down, down right and down left.

The intensity values are normalised so that the spread of pixels in each co-occurrence bin is optimised. This ensures a good textural representation when lesion intensity values are similar. Prior to the co-occurrence calculations, the normalised data is split into 4 bins (0-25, 26-50, 51-75, 76-100). In previous work the number of bins was varied between 4-32. It was found that the more sparsely populated bins did not improve the neural network results.

The divided data is converted into an occurrence matrix by comparing the pixel values next to each other. This process occurs for each direction of neighbouring pixels. The co-occurrence matrix is formed by dividing each member of the occurrence matrix by the total number of point pairs in the occurrence matrix. This new matrix known as a grey level co-occurrence matrix gives an estimate of the joint probability that a pair of points occurs. Several measures can be obtained from a grey level co-occurrence matrix, such as Energy, Correlation, Contrast, Homogeneity, Entropy.

$$Energy = \Sigma_i^{Ng}\Sigma_j^{Ng} p(i,j)^2 \tag{5.44}$$

$$Correlation = \frac{\Sigma_i^{Ng}\Sigma_j^{Ng}(ij)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{5.45}$$

$$Contrast = \Sigma_i^{Ng}\Sigma_j^{Ng}(i-j)^2 p(i,j) \tag{5.46}$$

$$Homogeneity = \Sigma_i^{Ng}\Sigma_j^{Ng}\frac{p(i,j)}{1+(i-j)^2} \tag{5.47}$$

$$Entropy = -\Sigma_i^{Ng}\Sigma_j^{Ng} p(i,j)\log(p(i,j)) \tag{5.48}$$

where Ng denotes the number of grey levels, $p_x(i)$ is the i^{th} row component of marginal probability, $p_x(j)$ is the j^{th} column component of marginal probability, μ_x is the average of p_x, μ_y is the average of p_y, σ_x is the standard deviation of p_x, and σ_y is the standard deviation of p_y.

5.3.4 Classification

The objective of lesion classification is to discriminate diabetic lesions from non diabetic lesions and image distractors such as light artifacts and retinal light reflexes. With bright and dark diabetic lesions having opposing characteristics and distractors, two pairs of classes are defined: exudates

versus bright distractors; and microaneurysms and haemorrhages versus dark distractors.

Given the number of classes, a rule is required to classify a candidate lesion into one of the predefined classes. This can be achieved through supervised learning in which features from observed candidates are used to direct classification procedures. To create a data set of observed candidate lesions, 10,000 bright and 10,000 dark regions from 1,000 random screening images were classified into the above classes by an ophthalmologist (see screening data section 8.4).

Supervised learning is achieved using a multi-layer perceptron (MLP) neural network, with back propagation and conjugate gradient descent training. A neural network was built for each of the two class pairs. Sensitivity analysis was performed on region descriptors to identify feature contribution - Five MLP neural networks were trained from five composite feature subsets for both bright and dark lesions:

1. shape and contrast
2. colour summary statistics and texture
3. green principal components and green summary statistics
4. Fourier principal components and Fourier summary statistics
5. finally Prewitt principal components and Prewitt summary statistics

MLP analysis for bright and dark subsets identified significant features forming two overall composite networks. The final networks consisted of 58 / 28 and 18 / 10 , for bright and dark lesions (features inputs / hidden units) respectively.

5.4 Evaluation

This evaluation is divided into two subsections: segmentation and classification. The performance of the region contrast gradient growth model is evaluated against four alternate techniques. Results from 100 sample lesions show that the proposed algorithm is over 200% more precise than the compared techniques in lesion segmentation. The performance of the lesion classification algorithms are generally presented in a per image or per eye basis and in conjunction with other retinal analysis components such as removal of ungradable images. This form of system evaluation is discussed

in detail in chapter 8. In this section, neural network evaluation via cross validation is discussed together with neural network performance, presented per lesion.

5.4.1 Lesion Segmentation

Benchmarking Method

The relatively small size of diabetic lesions in standard 760 × 570 fundus images can affect the accuracy of boundary demarcation. Absolute delimiting is therefore difficult to obtain. For this reason, the accuracy and precision of the presented segmentation algorithm is tested using 100 gold standard lesion boundaries obtained from 27 high-resolution fundus images with a 45 degree field of view, and photographed using a Canon fundus camera.

Lesions of varying sizes and contrast were selected to form a representative lesion collection, extracted from fundus images with different degrees of retinopathy. The dimensions of the high-resolution images were 3300 × 2600 pixels, approximately four times larger than the standard 760 × 570 lower resolution fundus used in lesion segmentation.

High resolution images were presented to an ophthalmologist in lesion centred 100 × 100 sub-windows. For evaluation convenience, each sub-window was centred on a discrete lesion. The ophthalmologist manually delimited the lesion boundaries by depicting each lesion edge pixel; followed by flood filling the periphery, creating a binary image. The high-resolution binary image down-sampled by a factor of 4 to test the algorithms at low resolution; a gold standard region of interest with sub-pixel accuracy is subsequently obtained. It is against this benchmark that the accuracy and precision of the algorithm is measured.

Models Tested

A comparison is made between the presented algorithm and four segmentation approaches - fuzzy C-Means clustering, recursive region growing, adaptive recursive region growing, and a colour discriminant function. All algorithms were implemented and evaluated against the gold standard region of interest.

The benchmark comparison with the aforementioned techniques was achieved by measuring the number of common pixels shared with the gold standard region of interest and the algorithm's

segmented area. For each gold standard region r, true positive tp_r, false negative fn_r, false positive fp_r and true negative tn_r metrics were calculated for each segmentation approach.

1. **True Positive** Number of correctly identified benchmark pixels.
2. **False Negative** Number of background pixels incorrectly identified as benchmark pixels.
3. **False Positive** Number of missed benchmark pixels.
4. **True Negative** Number of correctly identified background pixels.

The probability of the algorithm correctly segmenting a benchmark pixel and correctly identifying a background pixel are expressed in terms of sensitivity and specificity respectively.

$$Sensitivity = \frac{tp}{tp + fn} \tag{5.49}$$

$$Specificity = \frac{tn}{tn + fp} \tag{5.50}$$

Although sensitivity and specificity can demonstrate algorithm performance, technique comparisons based on single metrics are less ambiguous and therefore preferable. Accuracy is a commonly used metric that measures the probability of true positive and true negative classifications.

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn} \tag{5.51}$$

Unfortunately, the large ratio between the sub-window and a typical lesion size (26 : 1) prevents representative performance comparisons with this statistic. To illustrate this, consider, a benchmarked region R_1 and a test region R_2, housed within 51×51 sub-windows W_1 and W_2 respectively. Let the area contained within region R_1 equal 100 pixels. If test region R_2 segments only 10 of these pixels then the accuracy for R_2 is 0.965; see equation 5.52. This level of accuracy is due to the large number of true negative background pixels. When considered along side its sensitivity and specificity (0.1 and 1 respectively) the accuracy seems less informative, biased by the number of sub-window pixels.

$$Accuracy_{R2} = \frac{10 + 2511}{10 + 90 + 0 + 2511} \tag{5.52}$$

In an alternate approach, model precision is measured by summing the pixel deviation from the benchmark region. As segmentation precision is crucial for obtaining accurate region descriptors, false positive and false negative pixels can be deemed as equally damaging as they distort the regions shape. This deviation error metric is expressed in equation 5.53.

$$Error = FN + FP \tag{5.53}$$

Results

Analysing statistics for segmentation algorithms is a subjective process, balancing the importance of conservative and profuse growth. Pixel classifications (tp, fn, fp, fn) from each algorithm were returned for each of the 100 lesion sub-windows. The average sensitivity, specificity, accuracy and error for each algorithm is presented in table 5.1.

Table 5.1: Algorithm Performance Metrics

Model	Sensitivity	Specificity	Accuracy	Error
Contrast Gradient	97.18	98.05	98.58	36.69
Recursive	12.57	100	95.63	113.48
Fuzzy	88.36	93.87	93.99	156.16
Adaptive Recursive	91.19	92.13	92.45	196.15
Distriminant Colour	60.22	75.61	75.21	644.63

The statistical significance of observed relationship between algorithms as tabulated above (table 5.1) was verified to be significant. An ANOVA test was performed to test the null hypothesis that there are no mean differences between groups. The null hypothesis was rejected, implying that there is a significant difference between groups. As the ANOVA test specifies if there is a significant difference between groups and not which groups are significantly different to other, a t-test was performed to test the null hypothesis that the error mean of the Contrast Gradient model is equal to the next best model, the "Recursive" model.

The t-test value is -12.97. With 99 degrees of freedom in each sample, the critical value is -62.8004 at 99% confidence. The null hypothesis is therefore rejected, signifying that the Contrast Gradient model is better than the Recursive algorithm, even at the 99% confidence level. Similar pairwise comparisons indicate the Contrast Gradient model's superiority over the other alternative segmentation algorithms. In examining the pixel classifications for each algorithm (see table 5.2)

it is apparent that the other alternate algorithms either grossly under or over segment. With the Recursive region growing algorithm, the limited intensity threshold between seed and candidate region pixels is so small (10 pixels) that only partial segmentation is possible before the intensity difference reaches its threshold.

Table 5.2: Algorithm Pixel Classifications

Model	tp	fn	fp	tn
Contrast Gradient	126	3	35	2436
Recursive	15	113	0	2471
Fuzzy	114	15	141	2331
Adaptive Recursive	117	11	184	2286
Distriminant Colour	83	45	599	1872

The colour and intensity difference between lesion and background retina is commonly marginal with little contrast between the two. Consequently the fuzzy, adaptive and discriminant algorithms that do not utilise the limited edge strength, tend to overestimate the lesion boundary. This is particular the case when contrast is low and lesions are in close proximity. This is demonstrated in figure 5.8 where the Adaptive and Fuzzy means clustering algorithm segment two neighbouring lesions.

It is worth noting that the contrast gradient model makes on average 77 fewer fp and fn pixel misclassifications than next best performing algorithm (Recursive); that is an improvement of over 200%. Examples of contrast gradient segmentation can be seen in figure 5.10.

Effectiveness of Intensity

To evaluate the effect that different colour model components have on the algorithms region growing phase, the intensity component was compared to luminance and the green component.

Table 5.3: Effect of Colour Components in Region Growing

Colour Component	Error
intensity	36.61
luminance	36.82
green	36.69

The error metric difference between the aforementioned colour models components was found to be negligible and not statistically significant. Therefore, any of the three colour model components

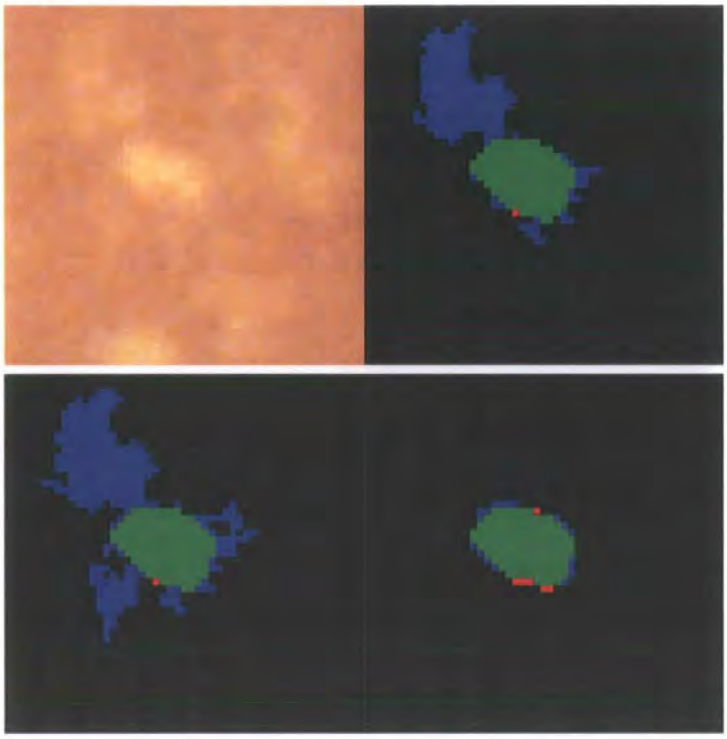


Figure 5.8: Segmentation Examples TL) Original, TR) Adaptive, BL) Fuzzy, BR) Gradient. Green TP, Red FN, Blue FP, Black TN

could be used in region growing and yield similar results; as depicted in table 5.3.

Effectiveness of the Gradient Image

The effectiveness of using a gradient image to locate a more precise boundary is evaluated against a standard intensity image, in which only the average intensity contrast between the grown region and boundary is measured. Table 5.4 shows the results of this test.

Table 5.4: Gradient Image versus Intensity Image

Model	Sensitivity	Specificity	Accuracy	Error
Gradient	97.18	98.05	98.58	36.69
Intensity	49.06	99.98	97.56	63.22

The results from table 5.5 demonstrate that the intensity model is more specific than the

Table 5.5: Gradient Image versus Intensity Image Pixel Classifications

Model	tp	fn	fp	tn
Gradient	126	3	35	2436
Intensity	68	61	3	2469

gradient model with less false positive classifications but is less sensitive, with a greater number of false negatives. This is shown in figure 5.9 where only the central lesion is segmented by the contrast intensity model. The precision advantage of using a gradient image over the intensity image is considerable with 43% fewer segmentation errors. Findings were statistically significant at 99% confidence.

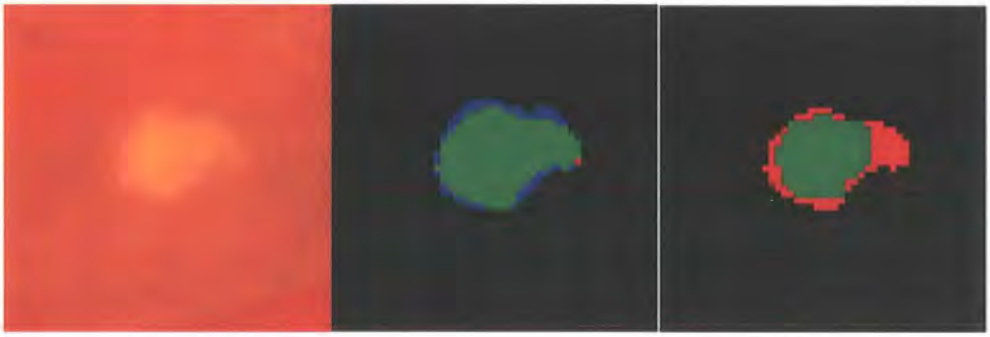


Figure 5.9: Gradient Image versus Intensity Image Segmentation Example L) Gradient, R) Intensity. Green TP, Red FN, Blue FP, Black TN

Effectiveness of the Stopping Criteria

The compactness of the stopping criteria is an integral part the proposed model. If the algorithm is stopped too early the grown region may not encompass the optimum boundary resulting in inferior segmentation. If the algorithm is stopped too late, the grown region may attach itself to a second lesion, thus creating uncharacteristic region descriptors and probably an incorrect lesion classification. The latter is common with the other approaches as seen in figure 5.8.

None of the 100 test images were prematurely stopped prior to reaching the ophthalmologists gold standard boundary. This compactness stopping criteria is preferable over an intensity thresh-

Chapter 6

Image Quality Assessment

6.1 Introduction

This chapter presents an algorithm for the automatic assessment of retinal image quality for lesion detection. The chapter focuses on the technical aspects of quantifying image quality into levels of gradability. Image quality is attained by summing a weighted contrast between vascular centre pixels and the background, multiplying the resultant by the contrast between the fovea and retina. It is assumed that blood vessel centrelines and the fovea have been previously located using the algorithms described in chapter 4. Images are subdivided into 5 categories of gradability. The algorithm is evaluated against three alternative approaches using a set of 200 fundus images from a normal screening population. The results demonstrate a significant performance improvement over other previously published approaches.

6.1.1 Motivation

The suitability of fundus images for diagnosis is an important issue in automated detection of diabetic retinopathy. Inadequate image quality can affect image classification as subtle visual differences between diabetic and non-diabetic lesions may become hazy leading to misclassification. In more extreme cases, loss of definition or obscurity may cause diabetic lesions to blend completely into the retinal background and be left unidentified and unclassified. To prevent unsound classifications, ungradable images require automatic exclusion from automated analysis and should

the contrast between the grown region and the region's boundary, identifying the optimum boundary where contrast is most significant. The compactness stopping criterion exploits the natural shape of lesions to halt the algorithm when sporadic growth occurs.

Experimental comparisons have been conducted on four segmentation approaches - fuzzy C-Means clustering, recursive region growing, adaptive recursive region growing, and a colour discriminant function. All algorithms were evaluated against a randomly-selected image set with ophthalmic lesion boundary demarcation, demonstrating the superiority of the proposed method.

Liberal feature extraction enabled comprehensive lesion representation using a number of shape and reflectivity descriptors. Neural networks have been employed for lesion classification, using sensitivity analysis to select suitable lesion descriptors that contribute to the network. K-fold cross validation was performed on the neural dataset, forming an ensemble of 10 networks. The large number of bright and dark image distractors caused imperfect neural classifications. Future work is therefore required to improve lesion representation and / or remove these distractors before classification.

are eventually used for training, selection and testing. The network performance error is estimated by average the performances of the individual networks. The individual network performance for both bright and dark cross-validation networks are depicted in figure 5.4.2; where $K = 10$.

Using the averaged performance measure as an estimate of the network performance on new data, the sensitivity, specificity and lesion classification for both bright and dark ensemble networks is shown in table 5.6. The performance figures indicate that the bright lesion classification is equally balanced between sensitivity and specificity whereas the dark lesion network is sensitivity biased; with under 2% false negative and 28% false positive lesion classifications. This relates to a large number of image distractors affecting the dark neural network.

Table 5.6: Exudate and Microaneurysms and Haemorrhages Ensemble Classification

Network	Sensitivity	Specificity	tp	fn	fp	tn
Bright	91.26	91.28	419	36	560	6676
Dark	98.06	72.8	152	3	4590	12273

During a further investigation into dark lesion image distractors, it was found that the majority of distractors occur from broken capillaries where the vessel segmentation algorithm failed to segment the full vascular structure.

5.5 Conclusion

Algorithms for the automated segmentation and classification of candidate lesions have been presented; an important stage in the prevention of sight threatening retinopathy. Although a number of algorithms have been published for lesion segmentation, many are unreliable due to marginal colour and intensity difference between diabetic lesions and background retina. This limited contrast has an adverse effect on alternate algorithms causing poor lesion boundary estimations.

In addition the intensity difference between the background retina and lesion seed can vary significantly across an image. Other region growing algorithms that use a fixed threshold rule for the inclusion of pixels are unreliable due to this variability.

In contrast, the presented segmentation algorithm iteratively selects the highest intensity pixel for growing regardless of the seed intensity. The algorithm also utilises the limited edge strength of candidate lesions by converting the fundal image from RGB to Gradient and iteratively compares

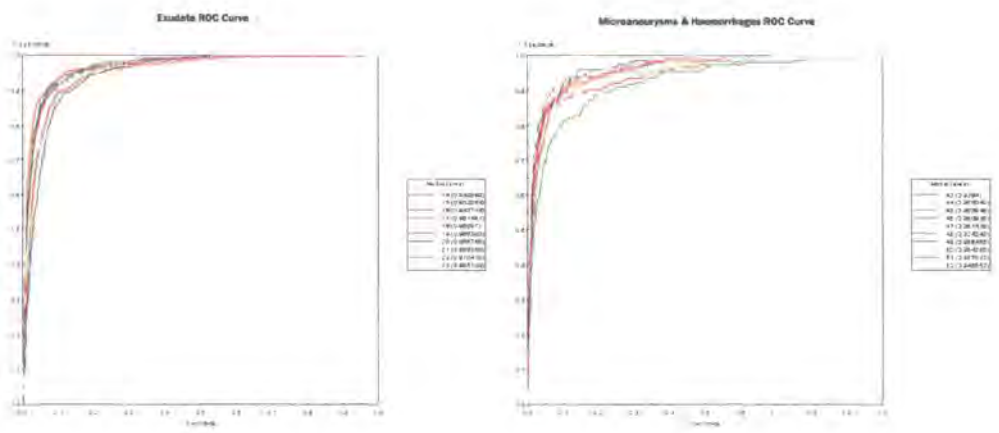


Figure 5.11: ROC Curves a) Exudate Ensemble Neural Network b) Microaneurysms and Haemorrhages Ensemble Neural Network

selection, testing) derived from a main dataset. Neural optimization is achieved using a training subset, while the selection subset is used to alleviate over learning by halting training. An unbiased estimation of the networks performance is then attained from the test subset. Unfortunately, these estimates may exhibit high variance.

A more reliable indicator of generalisation performance would be from averaging the performance predication from a number of test subsets that have been retrained. Unfortunately, the dataset of 1000 fundus images, contains limited examples of diabetic lesions within the macula; exudates were present in only 50 images whereas microaneurysms and haemorrhages were present in 73 images. This relates to a total of 459 bright and 155 dark macula lesions. Therefore not enough data exists to perform multiple runs with separate training, selection and test subsets.

Collecting a dataset large enough to satisfy these requirements was deemed logistically unfeasible due to the additional resources required for manual ophthalmic labeling, and the finite ophthalmic time available during this project.

This problem is overcome using K-fold cross validation, where the dataset is divided into K equal parts or folds. For each K experiment, $K-1$ folds are used for training and selection with the remaining fold for testing. The advantage of this approach is that all of the examples in the dataset

old approach which as seen by the recursive pixel classifications in table 5.2 frequently under segments.

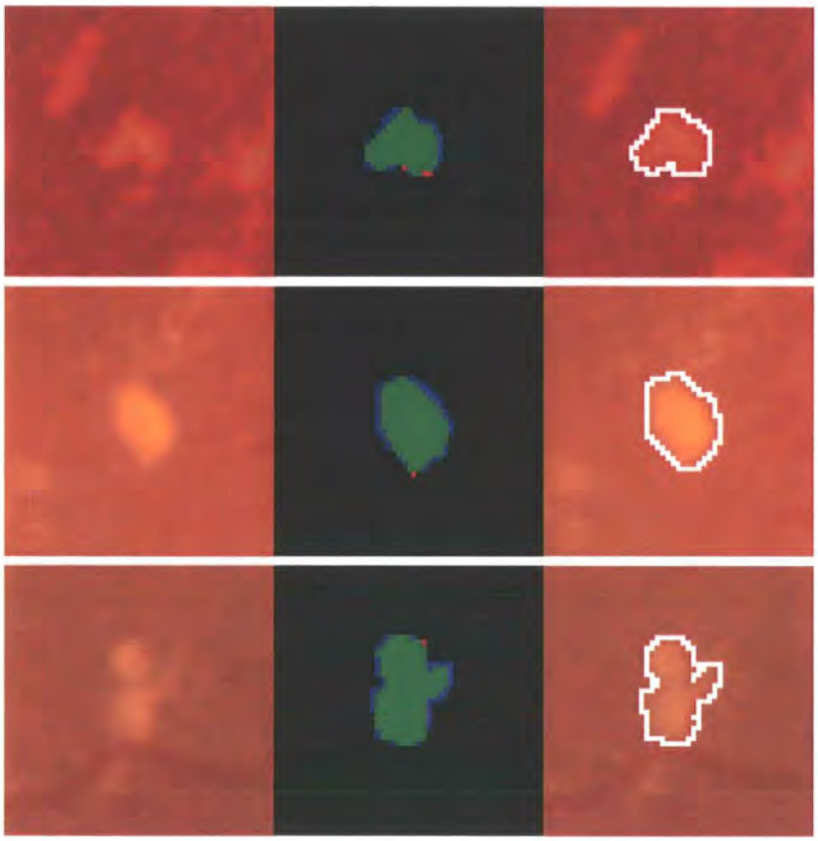


Figure 5.10: Contrast Gradient Model Segmentation Examples.

5.4.2 Neural Network

A problem with neural networks is the issue of generalization where the neural network begins to overfit the training data, meaning that it becomes fine-tuned to idiosyncrasies in the training data that will not necessarily appear in new datasets; subsequently determining the likely performance of new data is difficult.

Consequently, neural models are often built and assessed using three data subsets (training,

be flagged for either a repeat screening or ophthalmic review.

Defining image quality is a highly subjective abstract process based on the varying experiences, visual perception and the judgement of observers. In automated retinal analysis, a standardized so-called "typical experienced observer" is required to filter images into *achievable*, *minimum* and *ungradable* categories to avoid erroneous classifications from substandard images.

6.1.2 Chapter Contents

In section 6.2 the current definition of image clarity is discussed together with an overview of previously published image quality assessment algorithms. In section 6.3, the new quality assessment algorithm is described. The algorithm is evaluated against alternatives from the literature in section 6.4 and concluded in section 6.5.

6.2 Problem Domain & Literature Review

6.2.1 Image Quality Characteristics

Digital fundal photography allows instant image review and has quality assurance, yet poor quality images can still occur. Image quality can be effected by a number of factors including head or eye movement, poorly dilated pupils and/or a small pupil, blinking, and the presence of media opacity - most commonly caused by cataract.

Head or eye movement during photographic examination can degrade image acquisition resulting in out of focus, incorrectly illuminated or misaligned images. Retinal screening protocols require carefully aligned images to include defined areas of the retina. In a macula centred image, protocol requires the optic nerve head to be positioned in the midline, one disc diameter from the edge of the image field. Any movement just prior to acquisition can cause misalignment and to vital regions being excluded from the photograph. In addition to head or eye movement, poorly dilated pupils also affect image illumination creating dark low contrast images and can prevent lesion identification (see figure 6.3a).

Any media opacity (opacity of normally transparent eye tissue), that obstructs the field of view through the visual axis from cornea to retina can degrade the retinal image. The most common media opacity encountered in retinal screening is cataract. If fundal cameras capture retinal images

through cataract, images appear blurred and are often ungradable (see figure 6.3b). Images can also be obscured by eyelashes or the eyelid if blinking occurs during acquisition.

Currently, assessment of image quality is based on subjective interpretation of three definitions of image clarity (developed by Gloucestershire Diabetic Retinopathy Screening Programme [65] and adopted by The National Screening Committee (NSC) [66]). The general guidelines of image quality suggest using two 45° field images per eye. A macula centred field is essential and is referred to as the defined position. Optic disc and nasal fields are a bonus. Image quality is based on the macula centred images. The three defined levels are:

1. **Achievable standard:** Optic disc less than or equal to one disc diameter from the defined position. Small vessels clearly visible within one disc diameter of the fovea and optic disc and visible across more than 90% of remaining image(s); see figure 6.1.

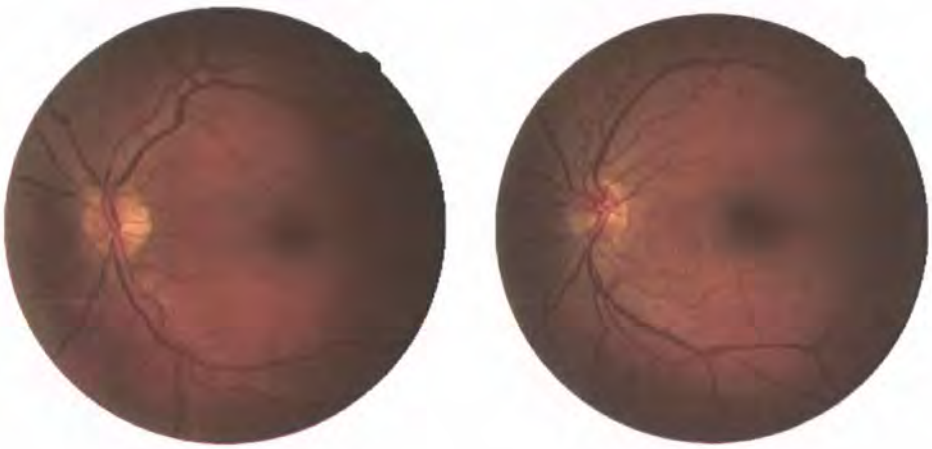


Figure 6.1: Achievable standard

2. **Minimum standard:** Optic disc less than or equal to two disc diameters from the defined position. Small vessels clearly visible within one disc diameter of fovea and optic disc and visible across more than 66% of remaining image(s); see figure 6.2.

3. **Inadequate (ungradable):** Optic disc less than or equal to two disc diameters from the defined position. Small vessels not clearly visible within one disc diameter of fovea and optic disc and visible across more than 33% of remaining image(s); see figure 6.3.

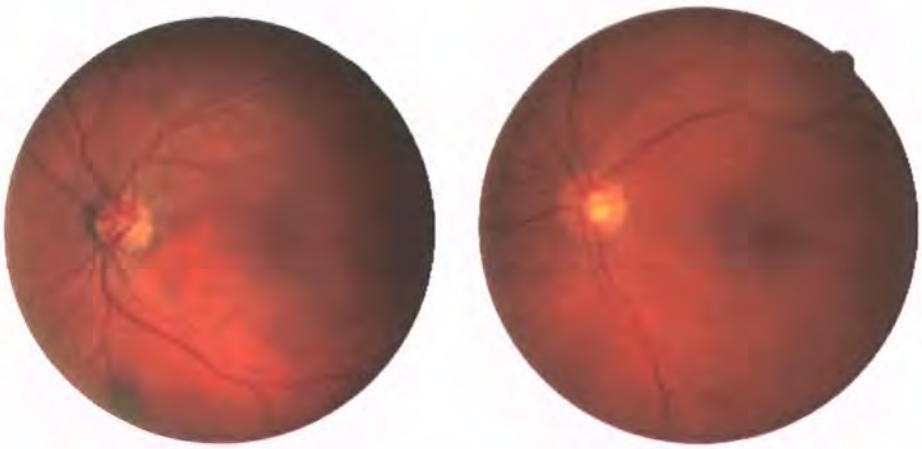


Figure 6.2: Minimum standard

6.2.2 Quality Assessment Algorithms

In order to comply with national screening guidelines, the image assessment algorithm should be capable of correctly partitioning images into the aforementioned three categories. To achieve this, a measure of quality is required; although image quality measures are well known in the domain of image restoration, diagnostic suitability is a relativity new research area with only limited publications. Usher *et al* [67] presented an algorithm to determine image quality using a quality metric based on the area of automatically detected blood vessel. Vascular segmentation was performed using a combination of orientated matched filtering and region growing, Usher report's a sensitivity of 81% and a specificity of 91%. The algorithm is described in detail by Himaga *et al* [68]. As blood vessels should be present in all retinal images regardless of ethnic origin or retinopathy, Usher measured vessel frequency from gradable and ungradable images to determine an image quality metric. Within each image an image quality metric score V was set from the total count of pixels classified as vessels. Images with blood vessel metrics above a threshold t_v were classified as gradable while images with metrics below t_v were classified as ungradable. The performance of this algorithm was evaluated using 800 images from 400 patients and comparing results to the opinion of three clinicians. An average inter-grader agreement was calculated using the average agreement between the system and the individual clinicians. The algorithm reportedly achieved 100% sensitivity and 94% specificity in detecting patients with at least one ungradable

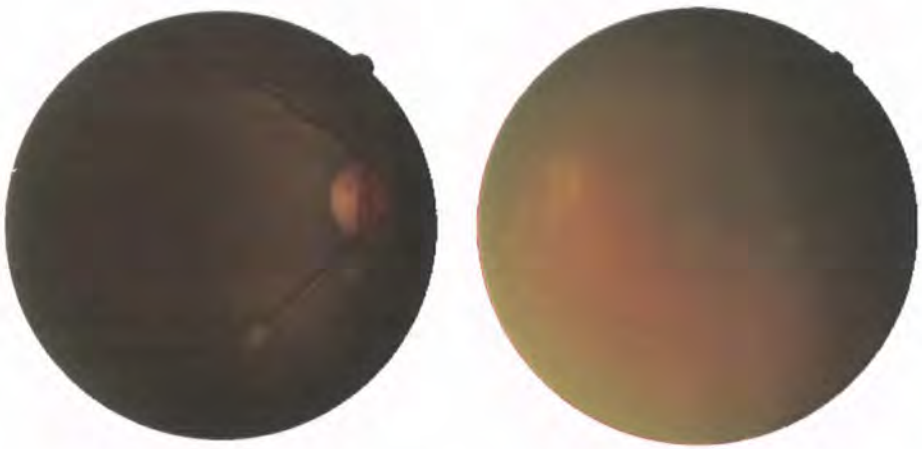


Figure 6.3: Inadequate (ungradable) a) poor illumination b) cataracts

image.

Lee *et al* [69] studied 360 retinal images from the Oklahoma native Americans and concluded that image quality could be defined by three parameters: brightness, contrast and signal-to-noise ratio (SNR). From the sample set, twenty images with excellent quality were selected. Quality parameters were obtained from these images together with an average intensity histogram - referred to as the desired values and the template intensity histogram respectively. Lee observed that the brightness, contrast and signal-to-noise values of an image were close to their respective desired values when the image's intensity histogram was close to the template intensity histogram and that these values could be derived from the histogram. An image quality measure was therefore proposed using the convolution of the template histogram with the image histogram and computing a quality index.

In an evaluation of Lee's work, Lalonde *et al* [70] examined the interdependency between image quality and histogram similarity in 40 retinal images of varying quality. During this study, histograms from several poor quality images were found to closely resemble the template histogram. In addition, histograms from several good quality images were notably different from the template histogram, signifying a weak connection between image quality and histogram similarity. Lalonde *et al* [70] experimented with distribution of edge magnitudes and the local distribution (as opposed to the global histogram of Lee) of pixel intensity as quality indicators. In a similar approach to

Lee, a typical edge magnitude histogram was formed using the edge maps from a set of good quality images. The difference between the typical and current image edge magnitude histogram formed a quality indicator. A second qualify indicator was derived by comparing local intensity distributions. This approach differs from Lee *et al*, by defining a set of local histogram templates instead of one global histogram template. Lalonde concluded that both quality indicators could help discriminate between good and bad images, although a larger image set was required to evaluate the performance of the approach.

6.2.3 Summary

As the guidelines defining fundus image quality are domain specific, it follows that automated techniques should aim to match the outlined criteria. Usher *et al*, came closest to achieving this but overlooked a number of important issues. First, blood vessels can be visible without being in focus, therefore as Usher's approach works on a thresholded blood vessel pixel count, blurred vessels can still add to this metric. Second, no emphasis is given to macula vessels. As macula vessel only count for a small percentage of the total vascular network, images with limited or no macula vessels could still be classified as gradable. Finally, the deviation from the defined position is not addressed. The image assessment algorithm proposed below takes inspiration from the National Screening Committee guidelines and addresses all of the these issues. The approach takes two independent measurements per image, calculating the contrast and quantity of visible blood vessels within one disc diameter of the fovea, and measuring the foveal contrast between fovea core and retina.

6.3 Method

The measurement of retinal image quality is concentrated in the macula region - two disc diameters around the centre of the fovea. This method assumes previous fovea localisation. It is also assumed that the vascular network has been previously segmented leaving a vascular map of segment centrelines (see chapter 4). The image quality assessment is performed on the green intensity channel of the colour RGB fundus images and is then normalised to increase its dynamic range.

6.3.1 Vascular measurement definition

The appearance of small blood vessels within one disc disc diameter of the fovea (macula) is the primary indicator of fundal image quality. There are three aspects of macula vessels that indicate quality - distance from fovea, contrast and quantity. The more blood vessels that are visible within the macula, the closer they are to the fovea centre and the higher the contrast between the vessels and the background retina, the better the image quality.

For each vascular segment, $i = \{1..S\}$, the average pixel distance away from the fovea and the contrast with the background retina is measured. The distance of a vascular segment i from the fovea is determined by averaging the distance between each pixel $p_0..p_{\eta_i}$ from centreline α_i and the fovea centre c; as expressed by equation 6.1. Let η_i represent the number of pixels contained within centreline α_i.

$$\overline{\gamma_i} = \frac{1}{\eta_i} \cdot \sum_{j=1}^{\eta_i} \sqrt{(p_{x_j} - c_x)^2 + (p_{y_j} - c_y)^2} \tag{6.1}$$

The contrast of vascular segment i is measured by taking the average intensity of centreline pixels α_i and subtracting it from the average intensity of boundary pixels β_i. To ensure that pixels measured by a segment's boundary are of the surrounding background retina and not the edge of its own blood vessel, boundary diameters are set to twice the average macula vessel diameter of 7 pixels. A segment boundary β_i is formed by dilating a segment centreline α_i by two structuring elements A and B; structuring element B is subsequently subtracted from A, where element $A = ones(6, 4)$ and $B = ones(4, 2)$. The mean pixel intensities of centreline α_i and boundary β_i are expressed by equation 6.2 and 6.3 respectively. Let $b_0..b_{\tau_i}$ represent the pixels within boundary β_i and I denote the normalised green channel of the fundus image.

$$\overline{\alpha_i} = \frac{1}{\eta_i} \cdot \sum_{j=1}^{\eta_i} I_{p_{x_j}, p_{y_j}} \tag{6.2}$$

$$\overline{\beta_i} = \frac{1}{\tau_i} \cdot \sum_{j=1}^{\tau_i} I_{b_{x_j}, b_{y_j}} \tag{6.3}$$

where τ_i represents the pixel count of boundary β_i. The intensity contrast for vessel segment i is calculated by simply subtracting $\overline{\alpha_i}$ from $\overline{\beta_i}$, as expressed by equation 6.4.

$$\omega_i = \overline{\beta_i} - \overline{\alpha_i} \tag{6.4}$$

In general, the greater the length of a visible macula blood vessel, the greater the quality. This quality improves further if the vessel has good contrast with the retina. The vascular metric for a vessel is therefore the product of the segment pixel length by its contrast measure. The final aspect of vascular quality is the distance away from the fovea; the closer a vessel is to the fovea centre the higher the quality and the smaller the segment's average distance. The vascular metric penalises distant blood vessels by dividing through by the average distance. The vascular metric φ_i for vessel segment i is expressed by equation 6.5. The overall vascular metric υ is simply the sum of individual vessel metrics; as expressed by equation 6.6.

$$\varphi_i = \frac{\eta_i \cdot \omega_i}{\overline{\gamma_i}} \tag{6.5}$$

$$\upsilon = \sum_{i=1}^{S} \varphi_i \tag{6.6}$$

6.3.2 Fovea Contrast Definition

The foveal contrast within the macula is the secondary indicator of fundal image quality, because its also indicative of lesion contrast. Macula regions with limited contrast between the fovea and background retina can cause lesion concealment, with lesions appearing either washed out or shaded depending upon the macula exposure. The foveal contrast is defined by comparing the core intensity of the fovea to the background retina of the macula. To compensate for any minor disparities between the real fovea centre and its estimated location, the foveal core is defined as a circular region with radius $r = 10$ (half that of an average fovea radius), originating at centre of the fovea. This reduced size ensures that intensity measurements are kept within the fovea. The fovea intensity $\overline{\varrho}$ is determined by taking the average intensity of each pixel $f_0..f_\varsigma$ from the foveal core, where ς is the core's pixel count; as expressed by 6.7.

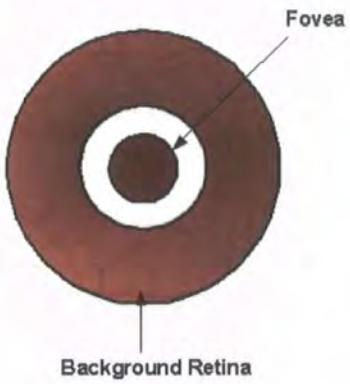


Figure 6.4: Fovea Quality Measure

$$\overline{\varrho} = \frac{1}{\varsigma} \cdot \sum_{i=1}^{\varsigma} I_{f_{x_i}, f_{y_i}} \tag{6.7}$$

To maintain maximum intensity differences between the fovea and background retina, a buffering zone is used to separate intensity regions; see Figure 6.4. The background intensity of the macula is measured in a region between two concentric circles originating at centre of the fovea with radius $r = 30$ and $r = 60$ respectively. The average intensity $\overline{\kappa}$ of each pixel $m_0..m_\epsilon$ within this region is defined as the macula intensity, as expressed by equation 6.8, where ϵ represents the macula pixel count.

$$\overline{\kappa} = \frac{1}{\epsilon} \cdot \sum_{i=1}^{\epsilon} I_{m_{x_i}, m_{y_i}} \tag{6.8}$$

The overall fovea contrast measure μ is simply the difference between the average intensity of the fovea core $\overline{\varrho}$ and the macula retina $\overline{\kappa}$; as depicted by equation 6.9.

$$\mu = \overline{\varrho} - \overline{\kappa} \tag{6.9}$$

6.3.3 Overall Image Quality Metric

An overall image quality metric χ, is defined by the product of the primary and secondary quality indicators. The vascular quality metric υ, incorporating the foveal distance, contrast and pixel quantity of each macula vessel segment is multiplied by the macula contrast μ between fovea and background retina; as expressed by equation 6.10:-

$$\chi = \upsilon \cdot \mu \tag{6.10}$$

6.3.4 Quality Metric Boundaries

The national screening guidelines for image quality divides images into three categories: *achievable*, *minimum* and *ungradable*. Deciding which category an image should join is a subjective process, and is especially so if an image is of a mixed quality lying between two categories. A wider categorisation of image quality was therefore deemed necessary to accommodate borderline images. A further two categories were added for images between *achievable* and *minimum*, and *minimum* and *ungradable* quality.

For the overall macula quality metric χ to be able to differentiate between image quality categories, metric boundaries are required for each quality division. This was achieved using a training set of 100 images that were assessed by an ophthalmologist, who categorised the images into 5 groups of ascending quality. Category metric boundaries were subsequently extracted; as shown in table 6.1.

Category	From	To
1	331	
2	101	330
3	36	100
4	6	35
5	0	5

Table 6.1: Quality Metric Boundary

6.4 Evaluation

The performance of the macula image quality assessment algorithm is evaluated against two alternate techniques. Results from 200 screening images showed that the macula model is 6% more accurate than the compared approaches and is capable of categorising image quality into 5 groups, matching the clinician's classification with an accuracy of 91% and detecting all clinically ungradable images.

6.4.1 Benchmarking Method

Fundus image quality guidelines specify that macula clarity must be such that blood vessel are clearly visible. With this in mind, 200 standard 760 × 570 fundus images were presented to an ophthalmologist, with the macula region of interest emphasized by an overlayed circle.

Images were graded by the ophthalmologist on a scale of $1-5$. Images with small blood vessels visible around the fovea and with good foveal contrast with the background macula area were graded as 1, images with similar vascular detail but with reduced foveal contrast were labelled as 2. Images that retained foveal contrast but only included macula periphery blood vessels where graded as 3, and were labelled as 4 with reduced foveal contrast. Where no vessels were visible, the image was labelled as 5. It is against this benchmark that the accuracy and precision of the algorithm is measured.

6.4.2 Models Tested

A comparison is made between the presented algorithm and two alternate fundal image assessment approaches - Lalonde *et al*'s [70] template intensity histogram and Usher *et al*'s [67] vascular metric. In Usher's algorithm, the vascular metric consisted of the sum of all pixels contained within the blood vessel network. In this study it was found more reliable to morphologically thin the segmented blood vessels to a centreline and sum the vascular centreline pixels. This reduces metric variability due to blood vessel width and treats all blood vessels with equal importance, whereas Usher's algorithm is biased by the pixels contained within the major temporal retinal vascular arcades (see figure 3.1 and macula vessels having little influence on the overall vascular metric. It is against this modified algorithm that the macula model is evaluated.

Lalonde stated that images could be crudely categorised into three groups: "good", "fair" and "bad". However, in testing Lalondes algorithm on 200 screening images, it was deemed impractical to perform a 3−way split due to the overlapping metrics within each category. Therefore, as the alternate algorithms can only distinguish between gradable and ungradable images, comparisons were made with benchmark classification $1-4$ deemed as gradable, and 5 ungradable.

6.4.3 Results

In this application it is important that the algorithm identifies 100% of ungradable images thus avoiding any potential misclassifications due to poor image quality. For this reason, performance is sensitivity biased, with metric thresholds dividing gradable and ungradable images selected from Receiver Operating Characteristic (ROC) curves where the sensitivity achieves 100%; see figure 6.5. The thresholds used to distinguish gradable and ungradable images for the macula model, Usher and Lalonde algorithms are 5, 4586.31 and 11419 respectively. Table 6.2 shows the results of the presented and alternate algorithms.

Usher [67] reported a sensitivity of 100% and specificity of 94%; however, in this evaluation, the specificity was 87%. Lalonde's algorithm achieves a poor specificity of 20% with 100% sensitivity, but results dramatically improve by allowing one false negative classification, giving 95% sensitivity and 81% specificity respectively. It is worth noting that the macula model algorithm has an accuracy of 94%, which is 6% more than Usher's algorithm and has almost half the false positive classifications.

Table 6.2: Image Quality Assessment Algorithm Performance

Model	Sensitivity	Specificity	Accuracy	TP	FN	TN	FP
Macula model	100	93	94	21	0	167	12
Usher	100	87	88	21	0	156	23
Lalonde	100	19.5	28	21	0	35	144

The observed relationship between compared algorithms was tested using an ANOVA statistical significance test, testing the null hypothesis that there are no mean differences between groups. The null hypothesis was rejected, implying that there is a significant difference between groups.

To test the null hypothesis that the error mean of the "macula model" is equal to the next best model, the "Usher's" model, a t-test was performed. The t-test value is 5.3355×10^{-025}. With

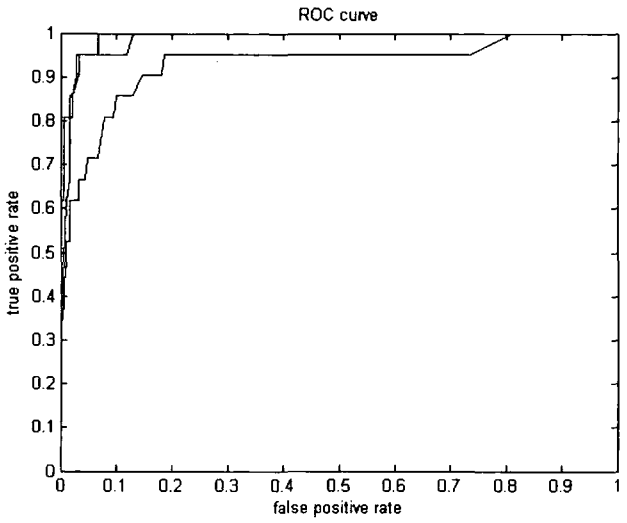


Figure 6.5: Image Quality Assessment Performance Receiver Operating Characteristic Curve: Macula model algorithm - Blue; Usher's algorithm - Red; Lalonde's algorithm - Green.

99 degrees of freedom in each sample, the critical value is 4.2207×10^3 at 99% confidence. The null hypothesis is therefore rejected, signifying that the Macula model is better than the Usher's image quality algorithm. Similar pairwise comparisons indicate the Combined vascular model's superiority over Lalonde's Template intensity histogram metric.

Alternate Amalgamated Approach

In an alternate approach, the Macula model is amalgamated with the amended version of Ushers vascular algorithm. In this new approach, the macula vascular contrast and foveal contrast measure is added to a macula excluded blood vessel centreline pixel count. In theory this should produce a more comprehensive image quality metric as the entire image is assessed.

Table 6.3: Amalgamated Image Quality Assessment Algorithm Performance

Model	Sensitivity	Specificity	Accuracy	TP	FN	TN	FP
Amalgamated	100	87	88	21	0	156	23

Table 6.3 shows that the amalgamated algorithm achieves the same performance as Usher's technique. Further investigation is therefore required to examine why there no performance ad-

vantages over the Macula model. One possible solution may be the inclusion of a contrast measure to the global vascular pixel count.

Image Quality Categorisation

The purpose of this algorithm is to detect ungradable images. The evaluation above considers clinically ungradable images. There is, however, a distinction between clinical and computerised gradability. To achieve an optimum classification it may be necessary to use a more stringent gradability criteria, thus ensuring only the sharpest images proceed to neural assessment. Although this ensures an idyllic classification environment, there is an increase in ungradable images which require manual review.

Table 6.4: Macula Model Image Quality Categorisation Performance

Grade	Correct	Incorrect	%
1	25	1	96
2	33	0	100
3	20	3	85
4	15	4	73
5	7	0	100
Total	92	8	92

The categorisation performance of the algorithms is evaluated against 100 of the 200 clinically assessed screening images. Table 6.4 shows that 91% of the automated image quality assessments matched the clinician. The remaining 9% are all within one grade of the clinician's classification. It is worth noting the system detected 100% of clinically ungradable images (grade 5). The high categorisation accuracy of this approach means that automated image quality assessment can not only exclude clinically ungradable images, but also exclude borderline cases leaving only the highest quality images for automated classification.

Effectiveness of foveal contrast measurement

In the presented algorithm, the foveal quality metric is determined by taking the average intensity difference between the fovea core and the background (macula) retina. To evaluate the effectiveness of using averaged measurements, an alternate approach is considered where an idealized foveal model is correlated (template matching) with the macula; see section 4.2.

Table 6.5: Foveal Measure Performance

Model	Correct	Incorrect	%
Foveal contrast	92	8	92
Foveal correlation	87	13	87

Table 6.5, shows that foveal contrast measurement produces 5 fewer classification errors fewer than template matching the macula with a foveal model. Findings were statistically significant at 99% confidence. With the characteristics of the foveal template incorporating a strong contrast difference between the fovea and the background retina together with an idealized , it was predicted that model correlation would yield superior results over averaged intensity differences. Further work is therefore required to examine the effects of alternate template models.

6.5 Conclusion

Automated assessment of fundal image quality is an important stage of automated retinal analysis, with inadequate image quality leading to unsound diabetic lesion and image classifications.

General guidelines for fundal image quality provide three definitions of image clarity; all referring to the visibility of small vessels within one disc diameter of the fovea. Previous image quality methods have concentrated on template histogram comparisons and global vascular pixel counting to identify ungradable images. The presented technique has focused on measuring the clarity of small macula blood vessels. Comparing the quality rating returned by the presented algorithm, Lalonde, and Usher, it was shown that the presented model is more accurate by 6%.

Using only macula vascular measurements, images could be correctly identified as gradable or non-gradable. By adding a foveal contrast measure, images can be further subclassified into 5 degrees of gradability. In addition to the further investigations into foveal template models and amalgamated vascular pixel metrics, further work is required to include optic disc clarity and image alignment as described in the National Screening Committee's (NSC) [66]) image quality guidelines.

Chapter 7

Retinal Vessels

7.1 Introduction

This chapter presents an algorithm for the automatic measurement of blood vessel diameters in retinal images. The chapter focuses on the specific technical issue of accurate diameter measurement, and does not address the other stages involved in the diagnosis of disease based on vascular pathology. It is assumed that approximate vessel center lines have been previously found. For further details see chapter 4. Diameter measurements are made by fitting a 2D model, which resembles an idealized cross sectional profile running along the length of a vessel segment in a small region of interest. The algorithm is evaluated against alternative approaches using a set of 100 cross section profiles from five down-sampled high definition fundus images. The results show that the algorithm is significantly superior to previously published methods.

7.1.1 Motivation

The retinal vessels are the only part of the central circulation that can be viewed directly and studied in detail [71]. As diabetes has a direct effect on blood vessels and changes in the morphology of the retinal vessels can be predictive of risk, segmentation and measurement of the retinal vessels is of central interest in automated retinal analysis.

A variety of morphological changes occur to retinal vessels in different disease conditions; however, this chapter is exclusively concerned with changes in the calibre (diameter, width) of vessels.

The change in width of retinal vessels within the fundus is believed to be indicative of the risk level of diabetic retinopathy [46]; venous beading (unusual variations in diameter along a vein) is one of the most powerful predictors of proliferate diabetic retinopathy [47].

Generalized and focal retinal arteriolar narrowing and arteriovenous nicking have been shown to be strongly associated with current and past hypertension reflecting the transient and persistent structural effects of elevated blood pressure on the retinal vascular network [72] [73]. In addition retinal arteriolar bifurcation diameter exponents have been shown to be changed significantly in patients with peripheral vascular disease and arteriosclerosis [74] and a variety of retinal microvascular abnormalities have been shown to be related to the risk of stroke [75]. Future precise retinal vascular analysis may allow risk stratification of a variety of arteriosclerotic diseases and clarify the microvascular contributions to clinical cardiovascular disease. It is worth noting that, in all these cases, it is not the absolute diameter of the vessel that is of interest, but variation in the diameter along a vessel.

Segmentation of vessels and measurement of the vascular diameter are two critical and challenging technical tasks in any system attempting automated diagnosis of vascular conditions. A number of methods for segmenting the vascular network have been reported in the literature [48] [76], with some success at least for larger vessels, although many papers gloss over the difficulty of reliably segmenting vessels in images of diseased retina, where there are significant distractors present; robust segmentation therefore remains an important research issue.

This chapter concentrates on the diameter measurement issue, whereas chapter 4 concentrates on the vascular segmentation issue. On a standard 30° fundal image with resolution 760 × 570, moderately sized vessels are as little as 6–8 pixels in width. Consequently, errors in diameter measurement of the order of a single pixel are large enough to significantly compromise attempts to characterize anomalies in vessel width, except for gross anomalies on the largest vessels. Even with improving fundal camera resolutions, it remains the case that there is a strong requirement for vascular diameter measurement algorithms with significantly sub-pixel sized accuracy.

7.1.2 Chapter Contents

In section 7.2 the retinal vessel structure and the true versus apparent width of the blood column are discussed together with an overview previously published diameter-measuring algorithms. In

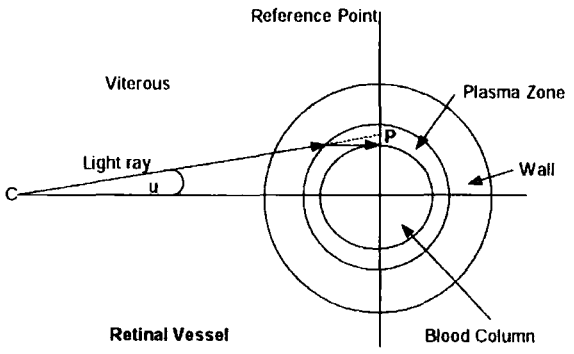


Figure 7.1: The apparent and true width of the blood column Based on Brinchmann-Hansen, 0; Heier, H. Acta Ophthalmolog. Suppl. 1986,179,29-32.

section 7.3, the new diameter-measurement algorithm is described. The algorithm is evaluated against alternatives from the literature in section 7.4 and concluded in section 7.5

7.2 Problem Domain & Literature Review

7.2.1 Retinal Vessel Structure

Light emitted from a fundus camera or ophthalmoscope is reflected back by the retina and choroids. Blood vessels absorb some of the reflected light, absorbing into the blood. When an image of a blood vessel is viewed from point C (see figure 7.1), 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.) [77]. The surrounding plasma zone and the vessel wall are transparent. Variations in the thickness of the vessel wall and the index of refraction have negligible influence on the apparent width of the blood column.

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 [77]. Looking at the retinal blood vessel as a vascular cross-section, the overall transmittance 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 has inspired many algorithms (see section 7.2.2). To complicate the problem further, some blood vessels include a light streak known as a light reflex that runs down the central length of the blood vessel. Naïve

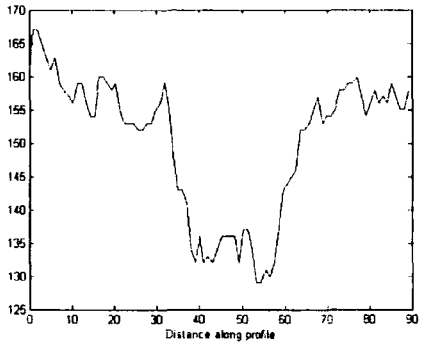


Figure 7.2: Sample vessel profile.

approaches can be fooled into thinking that the light reflex is the edge of the vessel.

The light reflex is understood to run across the surface of the plasma zone and the blood column and is believed to be generated from a rough reflecting surface and the intravascular column of erythrocytes [77]. Light reflexes are more common in younger retinas due to their increased reflective surfaces. However, arteriosclerosis can also affect the appearance of the light reflex by changing the reflective index of the vessel wall, thus increasing the observed intensity.

7.2.2 Diameter measurement algorithms

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. Direct methods that require ophthalmoscopic observations are often inaccurate due to non-standardized illumination, low magnification, and a subjective bias [78]. More objective measurements are possible from fundus images in which the film transmission has been converted to intensity [79] [80] [81]. 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 [82], so that the measuring "by eye" directly from green channel intensity images is difficult and error prone.

As a result, a vascular cross-section is often presented in graphical form; see figure 7.2. The intensity level of the background retina is often different on either side of the vessel and the vessel often contains a bright reflex.

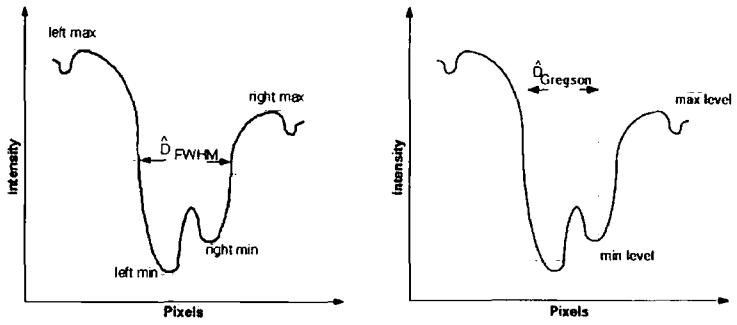


Figure 7.3: a) Full width half maximum. b) Rectangular profile.

Due to pulse variations in vessel calibre, retinal photographs taken at the same fundus position will not yield the same vascular width, if exposed at different times. However, the use of relative widths (i.e. the ratios of individual widths to the average vessel width) can ensure reproducibility, as pulse characteristics, as with variable illumination conditions from incident light, are discounted in ratios. It is also important to note that, for diagnostic purposes, obtaining consistent measurements is more important than obtaining the actual vessel widths.

Several previous authors have presented algorithms for measuring vascular diameters. Brinchmann-Hansen *et al* [82] [83] measured widths of retinal vessels using micro densitometry and observed the importance of the central light reflex which prevents naïve approaches from working well. They presented an algorithm called Full Width Half Maximum (FWHM) to measure vascular diameters. This approach calculates a "half height point" on the left and right sides of the initial estimated mid-point 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; see figure 7.3a. This approach is also called Half Height Full Width (HHFW).

Gregson *et al* [84] introduced an alternative approach, comprising of a rectangular profile of a fixed height that is fitted to the profile data. The height is fixed to the difference between the minimum and maximum intensity values in the profile. The width of the rectangular profile is adjusted until the area under the rectangular profile is equal to the area under the profile data;

see figure 7.3b.

Chapman *et al* [85], compares three methods of automated vascular measurements with manual recorded vessel diameters. A Gaussian function using non-linear regression, a Sobel edge detection algorithm and a sliding linear regression filter (SLRF) are all compared. Chapman states that using the Sobel edge detection algorithm to locate the blood vessel edges is unreliable due to a tendency to wrongly locate the central light reflex as the vessel edge. Chapman found the SLRF method to be most accurate and robust of these three methods; however, it is not practical for low-resolution images, as a minimum number of 10 pixels are required to estimate the slope by linear regression.

Chapman *et al* [85] and Zhou *et al* [86] have both experimented with Gaussian fitting functions to estimate vessel width. Chapman reported the Gaussian function fairing badly in comparison with the SLRF method. Zhou, on the other hand, reported promising results using a Gaussian model. Pedersen et al [46] presented an algorithm that used a cubic spline with 6 degrees of freedom to model the vascular profile and reduce noise. This method did not measure vessel diameters directly, but helped to reduce noise and relied upon Gregson's rectangular profile for width measurements. Gao *et al* [87] used a Difference-of-Gaussians to improve performance on vessels with a light reflex.

A number of authors have used matched filters to detect and segment blood vessels, including filters with Gaussian profiles. Gang *et al* [88] showed that the width control parameter of a gaussian profile matched-filter is linearly related to the actual blood vessel width. The majority of the aforementioned approaches use a one-dimensional cross-sectional profile to measure the vascular diameter. Matched filters, in contrast, form a two-dimensional profile that is more resilient to poorly positioned profiles, and poorly defined vascular edges, but are applied at set sizes (typically in steps of one pixel) [89].

This chapter introduces a two-dimensional model with a Gaussian or Difference-of-Gaussian profile, which is iteratively optimized to best fit the observed vessel. This allows a vessel width calculation to sub-pixel accuracy, and the smoothing introduced by the two-dimensional nature of the profile improves this accuracy. In this chapter the presented algorithm is compared to a Gaussian fitting function, HHFW and Gregson's rectangular profile and superior performance is demonstrated.

7.3 Method

7.3.1 Model definition

The 2D model in its most basic form consists of an idealized straight vessel segment, with a Gaussian profile, G_1, orthogonal to the vessel direction, which can be optimized to closely map a vessel segment; see figure 7.4. 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; variable-metric optimization is used to find a good set of parameters. As it is important that the model mimics the vessel segment being measured as accurately as possible, a variation can be used to model the light reflex that is apparent in some vessel segments. This is achieved by subtracting a second small Gaussian curve, G_2, from the main one. The second Gaussian is oriented at 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. As not all vessel segments have a light reflex, both models with and without light reflex are fitted and the model with the best fit is selected.

The 2D model without light reflex (see figure 7.4) has a profile shape governed by equations 7.1 and 7.2.

$$\alpha = x\sin\theta - y\cos\theta - \mu \tag{7.1}$$

$$f_{x,y} = t - h_1 e^{-s_1\alpha^2} \tag{7.2}$$

where x and y map the profile data range in this example from -10 to $+10$ in increments of 1. The model parameters are: t, the profile maximum; h_1, the height of the Gaussian; s_1, the width of the Gaussian, equal to $\frac{1}{\sigma_1^2}$; θ, the orientation of the model; and μ, the offset of the Gaussian center from the initial center estimate, orthogonal to the orientation. Figure 7.4 illustrates the model and parameters (barring μ).

The two-dimensional model with light reflex shares the same base parameters as its non light reflex counterpart. The light reflex variation to the core archetype consists of a Gaussian curve G_2 subtracted from G_1. Additional parameters are included to control the shape and size of the light reflex. These parameters, h_2 and s_2, set the height and width of the Gaussian G_2 respectively.

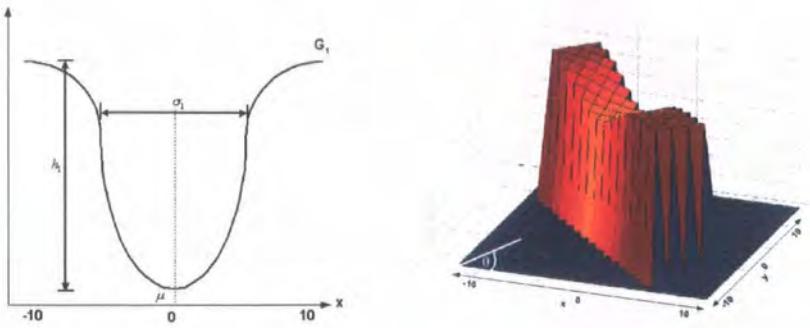


Figure 7.4: Model without Light Reflex. a) Cross-section of Profile. b) Two-dimensional Representation.

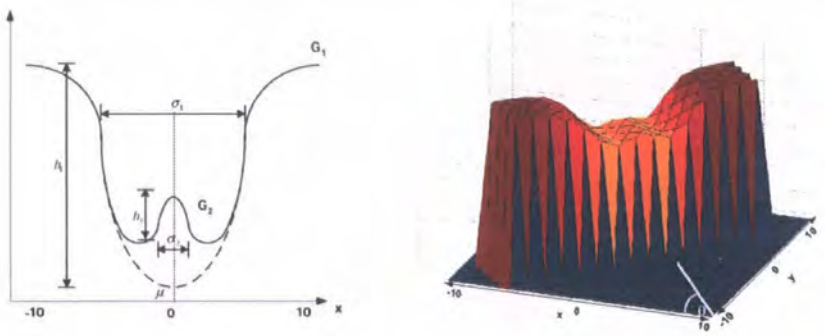


Figure 7.5: Model with Light Reflex. a) One-dimensional Representation. b) Two-dimensional Representation.

The model with light reflex is expressed by equation 7.3, and illustrated in figure 7.5.

$$f_{x,y} = t - (h_1 e^{-s_1 \alpha^2} - h_2 e^{-s_2 \alpha^2}) \tag{7.3}$$

7.3.2 Optimization

Two stage optimization process

A two-stage optimization process is used to orientate and shape the models to fit a 2D vessel segment, within an oriented rectangular region of interest centered at an initial point, (p_x, p_y), taken from the vascular profile, and oriented along the vessel line as estimated from previous and

next points in the vascular profile; see figure 7.6. In the first stage, a relatively long vascular region of interest is formed, the model width is fixed to a fairly broad value, and the algorithm determines the exact orientation, offset and height. In the second stage, a shorter region of interest is formed, the orientation is fixed, and the width, offset and height are determined to greater accuracy. Models both with and without a light reflex are optimized at each stage, the model which best fits the vessel segment is selected, and the width is calculated from the final parameters.

The two stage process is necessary for two reasons. First, the short model is poor at determining the angle as it lacks sufficient support along the vessel line. Second, it is prone to "default errors," where the model completely fails to capture the vessel shape (e.g. by forming a very wide, flat Gaussian fitted on one side to the background region) if the initial settings are too far from the correct values. The first stage, however, is very robust. It handles well the entire range of typical vessel widths, and inaccuracies in the initial vessel center line up to 66% (at a 95% confidence level there is no statistically significant change in vessel width measurements up to this disarticulation). In this research it is almost unknown for the algorithms that determine the initial line to exceed a 66% error in center line determination. Once the first stage has determined the angle, and reasonable starting values for the other parameters, the second stage determines a more accurate local width measurement, and is very reliable.

Vascular region of interest and mask

The 2D model is fitted to a local section of vessel, within a rectangular region of interest (ROI) that is oriented at the initial estimated angle, θ, of the blood vessel. The use of an oriented ROI is important as the background intensity varies fairly quickly, and there may be other vessel segments nearby; in early work using a square region of interest oriented with the image axes, it was found that such features in the corners of the ROI could have detrimental effects. For computational convenience a bounding rectangle of the ROI is calculated, and form a mask to identify the ROI; see figure 7.6. Calculations over the bounding rectangle are performed, and multiplied through by the mask, $z_{x,y}$, to ensure that only the pixels within the ROI take effect. Given a ROI of length L (along the vessel orientation) and width W (orthogonal to it), the mask is applied to a region area from $p_x - X$ to $p_x + X$ and $p_y - Y$ to $p_y + Y$, where $X = (L\cos\theta + W\sin\theta)$, $Y = (L\sin\theta + W\cos\theta)$.

The size of the ROI depends upon the stage of model optimization. In the first stage a 15×15

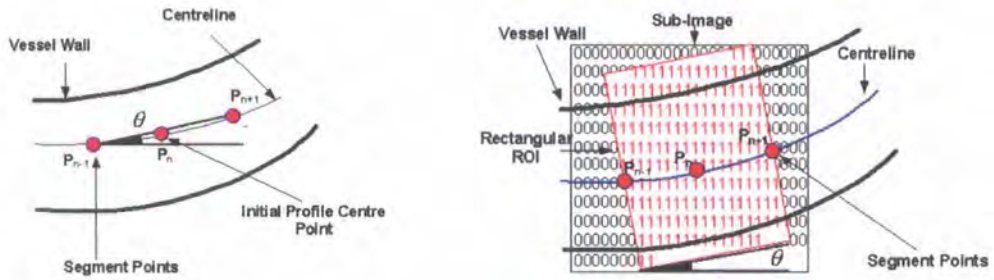


Figure 7.6: Vascular model. a) Vessel segments points. b) Mask and Region of Interest

ROI is used to enable the 2D model to fix the local angle against a sufficiently long vessel segment and to find the vessel "line" despite noise. In the second phase a 5×20 ROI allows the 2D model to obtain a good localized diameter measurement. With the average width of larger vessels in a 760×570 fundus images being $6 - 9$ pixels, both ROI are large enough to allow flexibility for off centered and beaded vessels, and to accommodate sufficient areas of the background.

Initial settings of model

The initial settings of the model parameters are extremely important. Bad initial settings may result in poor parameter optimization and an undesirable fit between model and vessel, leading to an inaccurate width measurement.

The initial settings for all of the parameters have been determined by trial and error. The initial parameter settings are $h_1 = 1.5d$, $h_2 = 0.5d$ where d is the difference between the minimum and maximum intensity values within the region of interest. The offset μ is set to zero. The width of the Gaussian G_1, s_1, is set to 0.1, giving a medium width with respect to the expected range, allowing the model to fit to both larger and smaller profiles. The width of the Gaussian G_2, s_2, is set to 0.8. The value of θ is determined by the vector connecting neighboring vessel segment points P_{n-1} to P_{n+1} (this gives a good enough starting value for the model to be optimized to the true angle).

Parameter Optimization

The optimization of model parameters is paramount to the algorithm's success or failure. Failure to minimize the error produces inaccurate profile widths. The noisy nature of the data can cause the optimization procedure to fail. In particular, the solution can diverge - for example, "running off" the side of the vessel and fitting a flattened Gaussian to the background. Carefully identifying initial parameter values, and performing a multiple stage optimization prevents these problems. In the first stage sensitive parameters such as the model angle, height and offset are fitted and a constrained "rough fit" is achieved; in the second stage all parameters except angle are optimized.

To effectively optimize the model, Quasi-Newton (variable metric) minimization [43] is used, a fast iterative algorithm that adjusts the model from the initial settings, by minimizing an error function that summarizes the goodness of fit of the current model. The procedure also requires a function giving the gradient of the error function with respect to each of the model parameters. The error function used is the sum-squared disparity between the model estimate and actual intensity values at the centers of the pixels in the region of interest, as expressed by equation 7.4.

$$E = \frac{1}{2} \sum_{x=-X}^{X} \sum_{y=-Y}^{Y} z_{x,y}(f_{x,y} - I_{p_x+x,p_y+y})^2 \tag{7.4}$$

where I is the image, f the model and z is the region of interest mask. The model is substituted by equations 7.2 and 7.3 for two-dimensional models without and with light reflex respectively. The differential of the error function 7.4 with respect to the model f is given by:

$$\frac{\delta E}{\delta f} = \sum_{x=-X}^{X} \sum_{y=-Y}^{Y} z_{x,y}(f_{x,y} - I_{p_x+x,p_y+y}) \tag{7.5}$$

Individual terms of the error gradient are derived by using the chain rule, multiplying equation 7.5 by the differentials of the model with respect to individual parameters. The partial differentials for the parameters h_1, h_2, s_1, s_2, μ and θ (see equations 7.1 and 7.3) are expressed below:

$$\frac{\delta f}{\delta h_1} = -e^{-s_1\alpha^2} \tag{7.6}$$

$$\frac{\delta f}{\delta h_2} = e^{-s_2\alpha^2} \tag{7.7}$$

$$\frac{\delta f}{\delta s_1} = h_1 \alpha^2 e^{-s_1 \alpha^2} \tag{7.8}$$

$$\frac{\delta f}{\delta s_2} = -h_2 \alpha^2 e^{-s_2 \alpha^2} \tag{7.9}$$

$$\frac{\delta f}{\delta \theta} = 2\alpha(-x \sin\theta + y \cos\theta)(h_1 s_1 e^{-s_1 \alpha^2} - h_2 s_2 e^{-s_2 \alpha^2}) \tag{7.10}$$

$$\frac{\delta f}{\delta \mu} = 2\alpha(-h_1 s_1 e^{-s_1 \alpha^2} + h_2 s_2 e^{-s_2 \alpha^2}) \tag{7.11}$$

After optimization the errors of the models with and without light reflex are compared. The model with the lowest error rate is selected.

7.3.3 Determining the model width

Once the model has been fitted, its parameters are used to estimate the vessel width. As previously noted, for further analysis it is important that the width measurement be consistent, rather than that it coincide with the user–perceived edge of the blood column.

For the single Gaussian model, the sigma parameter is the most appropriate estimator; this coincides with the inflection point on the slope of the vessel side. For the Difference-of-Gaussian model, the inflection point on the side slope is used; however, there is no simple analytical expression that yields its position. Instead, a simple line search procedure is used to locate a zero of the second derivative of the model, and the offset of this inflexion point defines the vessel width.

7.4 Evaluation

The performance of the method is evaluated against Brinchmann-Hansen's half height, Gregson's rectangular profile and Zhou's gaussian model. Results from 100 sample profiles show that the presented algorithm is over 30% more precise than the compared techniques and is accurate to a third of a pixel (a statistically significant at 95% confidence).

7.4.1 Benchmarking Method

The refractive index and the distance between the retina and the camera lens affect vascular diameters observed from a retinal camera [85]. Absolute measures of diameter are therefore difficult to obtain. The only appropriate method of testing and comparing a new measurement technique is to compare measurements from the same vessel. To this end the accuracy and precision of the presented algorithm is tested using 100 gold standard widths obtained from five high-resolution fundus images with a 45 degree field of view, and photographed using a Canon fundus camera. All 100-vessel widths were selected from non-tortuous vessel segments between bifurcations. The dimensions of the high-resolution images were 3300×2600 pixels, approximately four times larger than the standard 760×570 lower resolution fundus counterparts used in our screening programs. By manually measuring widths on the high-resolution images, and down-sampling the images to test the algorithms at low resolution, gold standard widths with sub-pixel accuracy are obtained. It is against this benchmark that the accuracy and precision of the algorithm was measured.

The width measurements from the aforementioned high-resolution images were taken at segment points and fixed angles derived by the algorithm during the blood vessel segmentation and profile data extraction phases. At segment point P the angle θ was used to plot an orthogonal profile line to the vessel centreline. The profile length from the full size fundus images was set to 80 pixels, 4 times larger than that used in the scaled image; the profile was constructed using bilinear interpolation.

A sophisticated technique called "kick points", due to Rassam *et al* [71], was used to manually calculate the vascular widths from the full size images. The "kick points" are visible points on the slopes of intensity profiles, defined as the first skew points. Kick points occur where the wall meets the most lateral extent of the blood column; the horizontal distance between the kick points indicates the blood column width (see figure 7.7) .

Rassam *et al* [71], states that the kick points approach is more accurate than finding the width at half the height; however, kick points are not always visible. If the blood column and the vessel wall have the same optical densities then kick points will not appear. To observe the kick points, the photograph must be high resolution and well focused, as the shape of the intensity curve changes with focusing. This makes the detection of the kick points difficult but has no effect on the degree of separation between points. In the test images, kick points were visible on the high-resolution

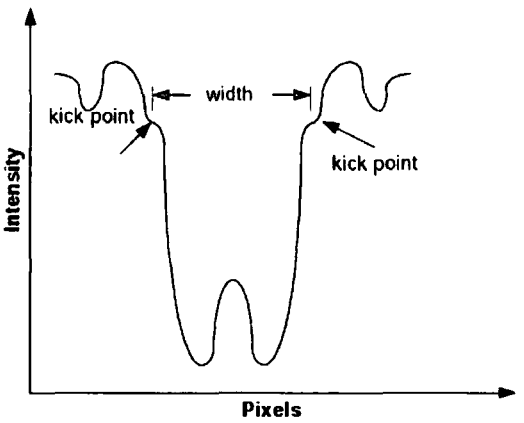


Figure 7.7: Kick Points

images, but not on the down-sampled versions.

7.4.2 Models Tested

A comparison was made between the presented algorithm and three other well-established approaches. The algorithms of Brinchmann-Hansen and Heier's full width half maximum [83], Gregson *et al*'s rectangular profile [84] and Zhou *et al*'s Gaussian model [86] were implemented using exactly the same intensity profiles calculated at the same segment points and angles as the 2D model. As previously discussed, the 2-D algorithm is given a rectangular region of interest at the initial profile center point. In contrast, the three benchmark algorithms use a one-dimensional profile which are determined by bi-linear interpolation along the profile direction.

One justification for using a 2-D model is that this smooths the local diameter estimates. As an additional comparison, the average width measurement from an ensemble of 1D profiles spanning the same range as the 2-D model was collected, to see if similar results can be obtained just by averaging 1-D profiles. The reasoning behind this study was to confirm that the 2-D model was not just behaving like a series of Gaussian models. In this test, if any of the ensemble profile parameters were outside a specified range (indicating a failure in model fitting) that profile was excluded from the averaging process (one-dimensional profiles are more susceptible to such instabilities than the 2-D model).

Table 7.1: Standard deviation of width difference at a scaling of 4

	σ_ϕ
2D Model	0.34
Avg 1D Gaussian	0.5
1D Gaussian	0.58
HHFM	0.62
Gregson	0.84

7.4.3 Results

In this application it is more important that the algorithm produces results that are precise (low error variance) than accurate (low mean error), as consistency is more significant than absolute diameter. Any consistent bias is easily subtracted to determine the true width. However if the results fluctuate, no compensation is possible. For this reason, the error standard deviation was used rather than the mean to assess the performance of the algorithms. The 100 width estimates ω_i returned from each algorithm were subtracted from 100 appropriately scaled benchmark widths ψ_i giving a width difference ϕ_i. The standard deviation of the width difference was then taken σ_ϕ.

$$\phi_i = \omega_i - \psi_i \tag{7.12}$$

$$\sigma_\phi = (\frac{1}{100}\sum_{i=1}^{100}(\phi_i - \overline{\phi})^2)^{\frac{1}{2}} \tag{7.13}$$

The difference in standard deviations between the 2D model and the alternative approaches was verified and found to be statistically significant. An F-test was performed to test the null hypothesis that the variance of the 2-D model is equal to that of the next best model, the "average 1-D" model. The F-test uses the variance ratio ($0.5^2/0.34^2$=2.212) to test for significant differences in variances. With 99 degrees of freedom in each sample, the F-test critical values are 1.39 at 95% confidence, 1.6 at 99% confidence. Consequently the null hypothesis is rejected, implying that the 2-D method is better than the average 1D model, even at the 99% confidence level. Similar pairwise comparisons indicate the superiority of the 2-D model over the other alternative algorithms.

Zhou *et al* [86] claimed that their approach, the 1-D Gaussian model, was more precise than other algorithms, including FWHM. This study supports their claim. However, the presented 2-D algorithm is 32% more precise again. The 1-D Gaussian model is also prone to failure to converge

sufficiently enough to obtain a sensible width reading (4% of diameters in our tests). Using an ensemble of 1D models eradicates this problem, and so the ensemble estimate out-performed the single Gaussian model.

It is worth noting that the 2-D model algorithm is accurate, on average, to 0.34 of a pixel; given that the manual process underlying the ground-truth measurements for the tests is accurate to only 0.25 pixels this is excellent performance. If a more accurate ground-truth were available it is possible that even better performance could be demonstrated (see figure 7.8a).

7.4.4 Effectiveness of the light reflex model

To assess the effectiveness of the model's light reflex in producing a more precise measurement, 20 benchmark widths that include a visible light reflex were used to compare against widths taken from the model with and without light reflex. It was found that the 2D model with light reflex gave results that were 13% more precise than given by its basic model counterpart in cases where the vessel segment included a light reflex. Although the improvement in the performance by using a light reflex model is marginal, it is statistically significant.

7.4.5 Effect of high curvature on precision

To evaluate the effect of high curvature vessel segments on the model's precision, 28 widths from high curvature points were calculated. It was found that at the 95% confidence level, there was no statistically significant change in width precision between normal and high curvature widths (see figure 7.9).

7.4.6 Effect of beaded vessels on precision

To assess the consequence of dramatic width changes within a vascular segment, 40 widths from beaded vessels were taken. The algorithm deteriorated, with the standard deviation increasing by 0.124 pixels to $\sigma_\phi = 0.464$ pixels (a statistically significant difference at 95% confidence). The other algorithms suffered a similar deterioration in precision, with Gregson's method performing second best at $\sigma_\phi = 0.771$ pixels. Comparing these two models show that the 2D model is better than Gregson's method and the findings were statistically significant at 99% confidence. (see figure 7.8b).

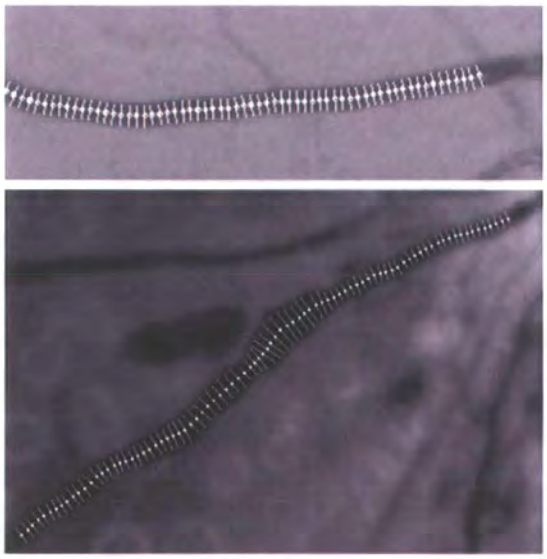


Figure 7.8: Detected widths on a) Straight b) beaded vessel segment.

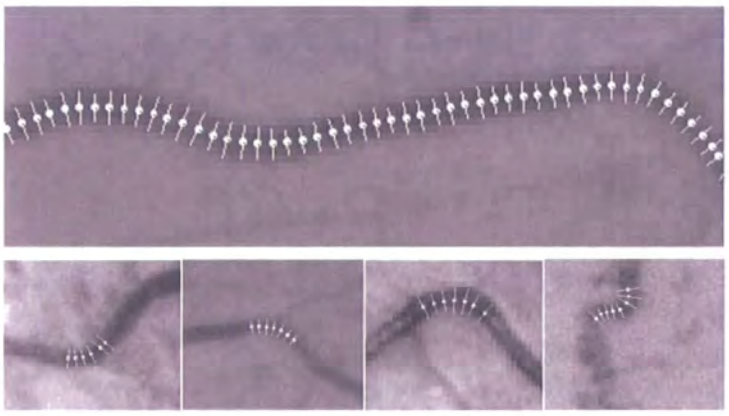


Figure 7.9: Detected widths on curved vessel segments.

7.5 Conclusion

An algorithm to automatically extract width measurements from retinal vessels has been presented. This plays a critical part in the structured analysis of the retina, and is potentially useful in the automated diagnosis of diabetic retinopathy.

Comparing the vascular widths returned by the presented algorithm, HHFW, Gregson and a 1-D Gaussian, it was shown that the presented model is more precise by over 30% (a statistically significant difference at 95% confidence). This is a considerable amount when small changes in the vascular diameter can alter the outcome of searches for pathology. During model fitting, a sophisticated optimization (Quasi-Newton) strategy was used, and appropriate energy function derivatives introduced.

The use of a 2-D model circumvents problems caused by noisy data such as small vessel branches and introduces a degree of smoothing that helps to improve the model fit. It was shown that the inclusion of a difference of gaussians model improves performance over a single gaussian where there is a visible light reflex. Finally, it is noted that the algorithm is robust enough to make sub-pixel accurate measurements on relatively low-resolution images, where vessels are only a few pixels wide.

Chapter 8

System Evaluation

8.1 Introduction

The focus of this research has been the development of a system to automatically detect early signs of diabetic maculopathy, with a long term aim to reduce the subjective classification variation and error of manual screening.

The progression of diabetic retinopathy and the related risk of vision loss can be clinically diagnosed by the appearance of a small number of diabetic lesions. Early signs of diabetic retinopathy may include out-pouching of weakened vessels walls (microaneurysms) and the leakage of blood (haemorrhages) and lipoproteins (exudates) from ruptured microaneurysms. The commonest cause of visual loss within the diabetic population is from diabetic maculopathy where diabetic lesions occur within one optic disc diameter of the central fovea (macula region). If the disease is caught and treated early in its progression, blindness can be prevented within 90% of patients at risk. Consequently diabetic patients should be annually screened for diabetic retinopathy. Screening may be carried out by a range of professionals with varying levels of subjective accuracy, including: general practitioners, nurse practitioners or ophthalmologists. However, even within this study disparities of 6.4% and 8.9% emerged between two clinicians classifying exudate and microaneurysms plus haemorrhages (HMA) respectively; see section 8.3.

There are two main applications for this system: a complete retinal photographic 'screening programme' in which the human grader is obsolete, automatically referring patients with ungrad-

able images or diabetic maculopathy to an ophthalmologist; and a 'pre-filtering' system in which patients with healthy retinae are removed from the screening images reviewed by human graders, thus reducing their work load.

Results from 1000 fundal screening images show that a sensitivity of 80% and specificity of 93% per patient for the detection of diabetic maculopathy as part of a 'screening programme'. For 'pre-filtering', a sensitivity of 100% and specificity of 92% for the detection of images with no diabetic maculopathy, reducing manual grading workload by 74%.

8.1.1 Chapter Contents

The evaluation sections of previous chapters have concentrated on comparing the presented algorithms to alternate techniques and previously published work. This chapter demonstrates the effectiveness of the overall system in identifying images with referable maculopathy and is assessed in a 'screening programme' and 'pre-filtering' role using two clinical HMA detection rules, and is evaluated against published results.

In section 8.2 the evaluation benchmark method is described, with the classification disparity between the two clinicians discussed in section 8.3. The screening data used is in section 8.4, and results shown in section 8.5. An evaluation summary concludes this chapter in section 8.6.

8.2 Benchmarking Method

According to the "Early Treatment Diabetic Retinopathy Study Group" [11] guidelines, sight threatening maculopathy is defined as patients with exudates or microaneurysms including haemorrhages within one disc diameter of the fovea; such patients are recommended for early referral.

Using this criteria images were manually classified (graded) as positive or negative for referable maculopathy, creating an image gold standard. For each image this manual classification consisted of four stages: macula demarcation, image quality assessment, exudate classification, and microaneurysm and haemorrhage classification.

In the first phase, the macula region was accurately identified by manually labelling the fovea centre and the boundary of the optic nerve head to determine the disc diameter (used to estimate the radius of the macula).

In phase two, using the national screening guidelines relating to image quality, images failing to meet the minimum criteria were labelled ungradable and excluded from further analysis.

For each image the system locates and classifies 100 potential exudate regions together with 100 potential microaneurysm and haemorrhage regions (see peak detection algorithm section 5.3.1). Phase three and four of the manual image classification therefore involved labelling 100 bright and 100 dark regions as exudates, microaneurysms, haemorrhages or other as appropriate. Image retinopathy grading was performed by two clinical research fellows with experience in diabetic retinopathy grading. A consultant ophthalmologist was used to adjudicate classification disparities between the two clinicians; see section 8.3. Any additional diabetic lesions not marked as a candidate lesion but identified by the clinicians were subsequently appended and labelled.

Although, at present, no formal detection criteria exists for referable maculopathy, it is generally agreed that any exudates within the macula are deemed as sight threatening. The significance of HMA's within the macula is a little more vague. For this reason, system performance is evaluated against two different detection criteria within one optic disk diameter of the fovea:

1. Any exudates and/or 3+ HMA's
2. Any exudates and/or 2+ HMA's

For the system to compare against the gold standard image classification, the system executes five retinal analysis components: optic nerve head segmentation, fovea localisation, image quality assessment, lesion detection and lesion classification. The high level of system connectivity means that poor accuracy and precision occurring within one component can have a knock on effect to subsequent components. For example, if the estimated fovea centre and disc diameter differ from the clinician's, then diabetic lesions present within the clinician's defined macula may not be present with the estimated macula region, causing a false negative classification; see figure 8.1.

To compensate for any localisation disparity between the system and the clinicians and to avoid false negative classification, the system can use conservative detection rules for HMA classifications. This conservative rule reduces HMA detection stringency by removing one HMA from the actual detection rule. Therefore if the real macula contains three HMA's and the system detects two it classifies the image as having maculopathy. This process is a double edged sword - although sensitivity increases, the number of false positives decreases specificity.

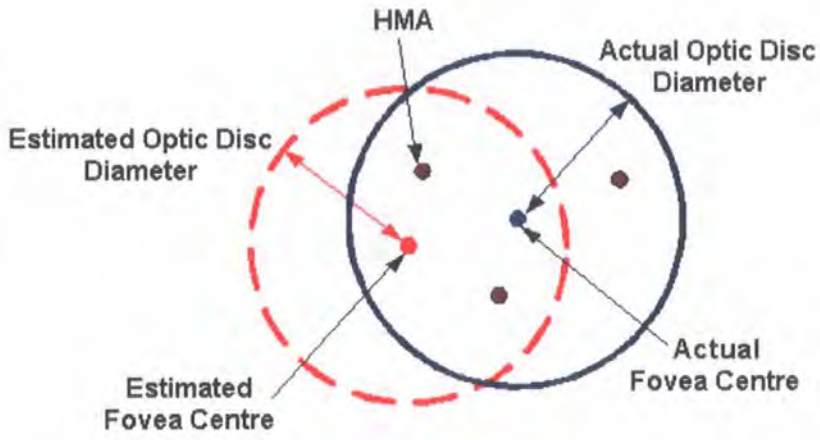


Figure 8.1: Potential Exudates missed due to estimated disc diameter and fovea centre.

A patient gold standard is defined from a patient's left and right image gold standards. A patient is classed as referable if either left or right fundal image is diagnosed with referable maculopathy. It is also worth noting that if an image from one patient's eye is healthy and the other eye is ungradable the patient is classed as ungradable; however, if the healthy eye is diagnosed as referable the patient is referable.

8.3 Clinician Disparity

A classification disagreement is defined as an image classified as containing retinal lesions by one clinician but not by the other. Even if both clinicians agreed that all but one candidates region were non-lesion, adjudication would be required on that image. However, if both clinicians classify multiple regions as lesions but disagree on a couple of suspect areas, the overall classification for that image would not be affected and therefore not require adjudication. The classification disparity between the two clinicians over 1000 images, labelling dark and bright candidate lesions separately is presented in table 8.1.

The 130 disputed images were reexamined by an adjudicating third clinician (consultant ophthalmologist) who stated that the images were difficult to classify due to borderline image quality. Results from reclassification (see table 8.2) show that no particular alignment existed between the adjudicator and the two clinicians; classification errors were therefore fairly evenly spread. This

Table 8.1: Clinician Disparity

Images	Bright	Dark
Gradable	914	914
Ungradable	86	86
Match	863	843
Disparity	55	75
Disparity %	6.37	8.89

suggests that it is genuinely hard to classify images with borderline image quality, resulting in subjective classifications. All of disputed images were included in the system evaluation after images were relabelled by the adjudicating clinician.

Table 8.2: Clinician Adjudication

Adjudicating Alignment	Bright	Dark	%
Clinician A	34	26	53.8
Clinician B	21	49	46.2

8.4 Screening Data

System performance was evaluated on 1000 retinal images obtained from a random screening population of 507 patients attending Birmingham City Hospital diabetic retinal screening program, the majority of patients (493) have paired left and right fundus images. Due to a plague of artificial eyes and image capture problems, 14 patients have only one fundal image. Images were acquired using a Canon CR6 45MNf fundus camera with a 45 degree field of view and resolution of 760×570. No data on age, type of diabetes, sex, ethnicity or duration were available for this study. However, Birmingham has a high population proportion of Asian origin, where Type 2 diabetes is prevalent.

In table 8.3, a breakdown of the screening images is shown. Of the 1000 screening images, 86 were classified ungradable due to obscurement or poor clarity of the macular region caused by cataract, poor illumination or badly focused photography. Out of the remaining 914 gradable images, the majority of retinae (76.8%) contained no abnormalities. The unbalanced ratio between normal and diseased retina means high levels of sensitivity are required to avoid a large false positive classifications from healthy retina, while being sensitive enough to identify sight threatening retinopathy. This problem is compounded by the number of image distractors such as the

Table 8.3: Screening Data

Per Image		n	%
Ungradable		86	8.6
Gradable		914	91.4
No Retinal Lesions		702	76.80
Background Retinopathy			
	Exudate only	26	2.84
	Microaneurysms/Haemorrhages only	173	18.92
	Combined	13	1.42
Maculopathy			
	Exudate only	35	3.8
	Microaneurysms/Haemorrhages only	58	6.3
	Combined	15	1.64
Drusen		152	16.63

152 images containing drusen, which are easily misclassified as exudate. It is also worth noting that a unbalanced ratio also exists between lesion types with a prevalence of microaneurysms and haemorrhages, outnumbering exudates by 65% within the macula. The importance of this imbalance however is debatable with ophthalmologists often comparing macula exudate significance to multiple microaneurysms and haemorrhages.

8.5 Results

System performance is evaluated in both 'screening programme' and 'pre-filtering' roles, with results presented per patient and per image against two HMA detection criteria. For each image, 100 bright and 100 dark candidate lesions were classified by a multilayer perceptron neural network ensemble with 10-fold cross validation (as described in section 5.4.2). The confidence levels produced by the network ensemble were averaged to give an overall classification prediction confidence. The activation level of the network that separates classes (lesion, non-lesion) was based on prediction confidence and was determined using a Receiver Operating Characteristic (ROC) curve.

8.5.1 Screening Programme

For the system to be assessed as a viable alternative to manual photographic grading, exceptional system performance is essential. With no human intervention between image classification and patient referral, any false negatives could result in patients at risk of visual loss, missing early

treatment, and too many false positive (and ungradable) ophthalmologist referrals cost valuable ophthalmic time and money. A balance between sensitivity and specificity is therefore required.

Although no formal standard exists, Diabetes UK have proposed that diabetic retinopathy screening programmes should achieve a sensitivity of 80% or higher, and specificity of 95% or higher. These guidelines for screening programme performance are against actual disease states which may or may not be identifiable in a digital image and are assessed on a per patient basis.

Without a full ophthalmic clinical examination on each patient, the true state of patient retinopathy is unknown. System performance is therefore evaluated against retinal photographic grading and not an actual patient retinopathy state.

All clinically ungradable patients and images were automatically removed prior to retinal analysis, and the "conservative" HMA detection rule was applied to both detection criteria, increasing the sensitivity while using a more stringent network threshold to increase specificity by reducing false positive classifications. During the removal of clinically ungradable images no false classifications were made.

Table 8.4: Screening Programme System Performance: Evaluated per Patient

Model	Sensitivity	Specificity	tp	fn	tn	fp	Clinically Ungradable	Referral
Exu ± HMA 3+	80	93	39	10	379	28	47	114
Exu ± HMA 2+	81	92	46	11	369	32	46	124

Table 8.5: Screening Programme System Performance: Evaluated per Image

	Sensitivity	Specificity	tp	fn	tn	fp	Clinically Ungradable
Exu 1+	74	96	37	13	838	26	86
HMA 3+	57	98	11	8	883	12	86
Combined	73	95	45	16	818	35	86
Exu 1+	68	97	34	16	843	21	86
HMA 2+	58	97	18	13	858	25	86
Combined	71	95	50	20	822	42	86

Working towards the sensitivity and specificity proposed by Diabetes UK's for screening programmes, and noting that the evaluation is against the patient's perceived retinal state, the system achieves a patient sensitivity of 80% and specificity of 93% for HMA 3+ and sensitivity 80%, specificity 92% for HMA 2+; see table 8.4. The difference in performance between the two HMA criteria is negligible, both per patient and per image; the breakdown of which shown in table 8.5.

With a specificity target of 95%, the number of patients without maculopathy (sight threatening retinopathy) being referred to an ophthalmologist was small (approximately 30). More concerning is the recommended sensitivity of 80% where $10 - 11$ patients with maculopathy are not being referred due to misclassification.

8.5.2 Pre-filtered Screening

Only a small proportion of retinal photographs reviewed by human graders containing diabetic abnormalities (23% within this screening set); consequently, the exclusion of healthy images could dramatically reduce grading workload and costs.

In pre-filtering screening, the automated exclusion of healthily diagnosed images requires 100% sensitivity, thus avoiding the misclassification of diseased retinal images. This is achieved through zero false negative classification's, however this is at the expense of higher false positive classifications.

Without pre-filtering, retinal graders are required to classify all screening images (1000 images in this research). Therefore even safely identifying and discarding 25% of the 702 unsystematic images, the human grader's would be presented with 175 less images.

Excluding Clinically Ungradable Images

To evaluate the system's performance as a pre-filtering screening tool, clinically ungradable images are first automatically removed from the screening set, and the neural network threshold levels are set to emphasise sensitivity. Using two HMA detection rules, system performance for both per patient and per image are shown in table 8.6 and table 8.7 respectively.

Table 8.6: Pre-filtered Screening System Performance: Evaluated per Patient

Model	Sensitivity	Specificity	tp	fn	tn	fp	Clinically Ungradable	Screening
Exu ± HMA 3+	97	80	48	1	320	78	47	173
Exu ± HMA 2+	96	67	55	2	264	128	46	229

The difference in false positive HMA classifications between the different detection rules as shown in table 8.6 and table 8.7, cause a significant specificity disparity of $13-14\%$ for both patient and image evaluations. This disparity affects the number of images requiring human grading, with

Table 8.7: Pre-filtered Screening System Performance: Evaluated per Image

	Sensitivity	Specificity	tp	fn	tn	fp	Clinically Ungradable	Screening
Exu 1+	90	92	45	5	795	69	86	
HMA 3+	94	92	18	1	827	68	86	
Combined	95	86	58	3	737	116	86	260
Exu 1+	90	92	45	5	795	69	86	
HMA 2+	90	78	28	3	689	194	86	
Combined	95	77	67	3	653	194	86	347

260 images requiring manual screening using the 3+ HMA rule and 347 with the 2+ HMA rule. These screening figures are determined by summing the number of ungradable, false positive and true positive images; see table 8.7.

Using the more stringent of the two grading criteria (Exu $\pm$ HMA 3+) the system achieves sensitivity of 97% and specificity of 80% (per patient), making one false negative classification; caused by a single faint exudate in a medium quality image (quality grade 3).

Comparisons with other automated referable maculopathy algorithms is restricted due to limited publications. Usher *et al.* [90] is the only truly comparable approach. Using an adaptive recursive region growing algorithm, Usher *et al.* [90] extracted an unspecified number of features into a multilayer perceptron neural network. Measured on a per patient and image basis, and using both HMA classification criteria, Usher's results, are inferior to the pre-filter system, with a performance advantage over Usher's algorithm of $1 - 15\%$ sensitivity and $10 - 15\%$ specificity; depending upon the decision rule and whether detection is performed per patient or image.

Table 8.8: Pre-filtered Screening vs Usher: Evaluated per Patient

Model	Criteria	Sensitivity	Specificity
Usher	(Exu $\pm$ HMA 3+)	88	62
Pre-filter	(Exu $\pm$ HMA 3+)	97	80
Usher	(Exu $\pm$ HMA 2+)	95	52
Pre-filter	(Exu $\pm$ HMA 2+)	96	67

Excluding System Ungradable Images

To achieve 100% sensitivity and prevent false negative classifications, only high quality retinal images may be assessed by the system. Restricting automated analysis to high quality images

Table 8.9: Pre-filtered Screening vs Usher: Evaluated per image

Model	Criteria	Sensitivity	Specificity
Usher	(Exu ± HMA 3+)	80	75
Pre-filter	(Exu ± HMA 3+)	95	86
Usher	(Exu ± HMA 2+)	87	67
Pre-filter	(Exu ± HMA 2+)	95	77

(quality grade 1-2), allows images with no visible diabetic abnormalities to be reliably excluded from human grading with 100% confidence. The negative side of this restricted analysis is the increase in ungradable images requiring human grading. Table 8.10 shows that the system classified an additional 106 and 102 patients (HMA 3+ and HMA 2+ respectively) above the clinical classification as being ungradable.

Table 8.10: Pre-filtered Screening System Performance: Evaluated per Patient

Model	Sensitivity	Specificity	tp	fn	tn	fp	System Ungradable	Screening
Exu ± HMA 3+	100	95	27	0	298	15	153	195
Exu ± HMA 2+	100	91	33	0	284	28	148	209

Table 8.11: Pre-filtered Screening System Performance: Evaluated per Image

	Sensitivity	Specificity	tp	fn	tn	fp	System Ungradable	Screening
Exu 1+	100	94	25	0	763	42	170	
HMA 3+	100	95	10	0	795	25	170	
Combined	100	92	31	0	743	56	170	257
Exu 1+	100	94	25	0	763	43	170	
HMA 2+	100	87	17	0	594	82	170	
Combined	100	87	37	0	695	98	170	305

Table 8.11 shows the system performance and the number of images requiring manual screening for each of the two HMA detection rules. The table also shows that the system flagged an additional 84 images above the clinical classification as being ungradable. The best results occurred from the 3+ HMA rule, in which 257 from the 1000 screening images required manual review, and 743 images with no sight threatening retinopathy (according to the detection rule) were automatically excluded. In terms of manual grading, this is workload reduction of 74.3%. Relating the pre-filtering to patents, table 8.11 shows the system's performance is identifying patients whose retinal photographs require manual grading. Both detection criteria reduce the number of patients

requiring further review by over 58%.

8.6 Summary

The role of the system as a 'screening programme' and for 'pre-filtering' purposes have been assessed, with system performance evaluated for each application against a gold standard data set consisting of 1000 randomly screened fundal images from 507 patients. For each image, two ophthalmic clinicians manually labelled diabetic lesions, the optic disc, the fovea centre and image quality, with a third clinician acting as adjudicator.

The system performance as a 'screening programme' achieved a sensitivity and specificity of 80%, 93% respectively; 2% under Diabetes UK proposed specificity guidelines. It is worth noting that the presented figures are against retinal photographic grading and not an actual patient retinopathy state. Therefore actual system performance maybe lower. Used as a automated screening tool, $10 - 11$ patients with sight-threatening maculopathy were misclassified.

For 'pre-filtering', a sensitivity of 100% was used to automatically exclude healthy images (images without referable maculopathy) and avoid misclassifying diseased retinal images. To achieve 100% sensitivity, retinal analysis was restricted to high quality images (quality grade 1-2). The maximum specificity at this level was 92%; reducing the workload for human grading by 74.3%.

Chapter 9

Conclusion

9.1 Introduction

The aims of the research presented in this thesis have been to investigate and develop techniques for a structured analysis of the retina, that singly analyse retinal features and collectively detect sight-threatening retinopathy (maculopathy). This is achieved amongst the high variability of the retinal structures, light artifacts, noise, diabetic and none diabetic lesions.

Several new and adapted algorithms have been specifically designed to accurately locate, segment and classify retinal structures together with diabetic lesions from the aforementioned image distractors. It is this development that has formed the basis of the work presented in this thesis.

9.1.1 Chapter Contents

The performance success together with the novelty of the developed algorithms are discussed separately. In section 9.2 the success of the developed algorithms are described in relation to their performance when evaluated against clinician labelled gold standards. In section 9.3 the research contribution or novelty of the developed algorithms are described together with a comparison overview against alternate algorithms. This is followed by future work in section 9.4, and section 9.5 concludes this chapter and thesis.

9.2 Criteria for Success

The criteria for success as set out in chapter 1, identified five areas of research where high algorithm performance was essential for automated retinal analysis and the early diagnosis of sight-threatening retinopathy. The performance of each criteria is examined in turn, with its effect on the overall system discussed.

1. **Locate and segment the Optic Nerve Head from the fundus image.**

   An optic disc inspired model located 90.7% of optic centers within 15 pixels; 8.3% within 25 pixels; and 1% over 25 pixels. The accuracy of the optic model was sufficient for initialisation of the optic disc deformable model.

   The demarcation results achieved from the deformable model when compared to an the ophthalmic gold standard, attained 73% good image segmentations, 10% fair and 17% poor. Although the algorithm can improve, the majority of optic nerve head boundaries were fully segmented, preventing false positive exudate classifications on the optic boundary and enabling approximate fovea localisation.

2. **Segment the boundary of potential retinal lesions indicative of diabetic retinopathy.**

   Lesion segmentation was achieved using a contrast and gradient based region growing model. Results from 100 sample lesions showed that the contrast and gradient model achieved an accuracy of 98.58% when compared to 100 clinician delimited lesion boundaries. The attained level of segmentation accuracy enabled accurate classification of potential lesion, as reliable neural classifications depend directly upon regional descriptors obtained from precise boundary segmentation.

3. **Classify potential lesions into retinopathy and non-retinopathy categories.**

   An ensemble neural network successfully distinguished between bright/dark lesions and their respective image distractors with high percentages of accuracy. Measured on a per lesion basis, the sensitivity for bright/dark lesions classification was 91.25% and 98.06% respectively with specificity reaching 91.28% and 72.8%. Although the presented network performance is acceptable, ideally the network sensitivity and specificity for both bright and dark lesions

should be over 90% and preferably even higher - as one false candidate lesion classification can alter an overall image classification.

4. **Determine the image quality and assess gradability of image.**

   The suitableness of fundus images for automated retinal analysis was identified using an image quality assessment algorithm. Comparing against 200 clinician graded images, the approach correctly identified 100% of non-gradable fundus images, together with subclassifying images into degrees of gradability; achieving an accuracy of 91% when compared to the clinicians quality assessment. The accuracy of the image quality algorithm allowed images with inadequate quality to be automatically identified, preventing unsound diabetic lesion and image classifications.

5. **Vascular width measurements**

   Changes within the vascular width are believed to be indicative of diabetic retinopathy risk [46]. Unusual width variations called venous beading are believed to be a powerful predictor of proliferate diabetic retinopathy [47]. Diameter measurements made by fitting a 2D model to 100 vessel segments, showed that the algorithm is accurate, on average, to 0.34 of a pixel. When small changes in the vascular diameter can alter the outcome of searches for pathology, such accuracy is significant.

6. **Classify image into referable and non-referable maculopathy.**

   The success of the image classification depends greatly upon the detection criteria and the role of the system. However, the maximum sensitivity and specificity for automatically detecting referable maculopathy was 100% and 92% respectively. This was achieved by restricting automated analysis to the high quality images (quality grade 1-2).

9.3 Summary of Research Contributions

This thesis has presented several pertinent research contributions in pursuit of automated retinal analysis. The primary contributions are summarised as:

1. **Chapter 3: Development of localisation model and deformable model for optic nerve head segmentation.**

Initialised by a localisation algorithm, optic nerve head segmentation utilised active contour techniques that had not been extensively examined in this domain. Novel additions to a deformable model exploited the disc's natural shape and used the disc's gradient characteristics and temporal locking to avoid images distractors such as blood vessels and the pallor.

The performance of the temporal lock deformable model was compared against a Hough transform and alternate de-vascularization variations. In combination with de-vascularization and the vector gradient, an algorithm ("DV-Hough") emerged with performance close to that of the temporal lock algorithm. However, with the temporal lock achieving better results in 59/90 cases, the superiority of the temporal lock algorithm was demonstrated [26].

2. **Chapter 5: Development of contrast gradient region growing algorithm with compactness stopping criteria.**

   A region growing algorithm using traditional intensity pixel aggregation region growth was amended to include a novel technique to identify the optimum region boundary; iterately calculating the average gradient contrast between boundary and current region. The growth algorithm was terminated when region compactness exceeded a threshold; selecting the optimum boundary at the growth point of maximum gradient contrast.

   A comparison was made between the presented contrast gradient region growing algorithm and four established segmentation approaches - fuzzy C-Means clustering, recursive region growing, adaptive recursive region growing, and a colour discriminant function. The contrast gradient model makes 68% fewer false positive and false negative pixel misclassifications than next best performing algorithm.

3. **Chapter 6: Development of fundus image quality metric.**

   A novel image quality metric was derived by multiplying cumulative weighted macula vessels contrasts with an overall macula contrast. Evaluated against the alternate algorithms of Lalonde *et al*'s [70] and Usher *et al*'s [67], the presented macula model nearly halved the false positive image classifications of the next best performing technique, with a specificity advantage of 6%.

4. **Chapter 7: Development of vascular segment model for vessel width measure-**

ments.

A flexible (novel) 2D model was designed to closely map a vessel segment. With model size and shape controlled by a small number of parameters, variable-metric optimization adjusted the model to fit the actual vessel profile. Once fitted, the model's parameters were used to estimate the vessel width.

A comparison was made between the presented 2D model and three other well-established vessel measuring approaches: Brinchmann-Hansen and Heier's full width half maximum (HHFW) [83], Gregson *et al*'s rectangular profile [84] and Zhou *et al*'s Gaussian model [86]. Comparing the vascular widths returned by the presented 2D model and the algorithms of Brinchmann-Hansen and Heier, Gregson and Zhou to clinician labelled vascular profiles showed that the presented model is more precise by over 30% [89].

9.4 Future Work

Due to complexities and variability of diabetic retinopathy, there are many directions in which this research could progress. The modular approach used within this research (see figure 1.4), means that automated detection of diabetic maculopathy can continually improve as additional layers of retinal understanding are added to the current system. Using this research as a basis, future work could enable the identification of other indicative clinical features that appear during different phases of progression and with varying levels of regularity. Also, investigating lesion identification at different image resolutions could allow earlier detection of currently identified lesions. Some of the earliest indictors of diabetic retinopathy can be identified from vascular changes, however, complete (or near complete) vascular segmentation is still elusive. To improve the system performance, these areas require further investigation.

1. **Vascular Segmentation**

   Further investigation is required to exploit the presented vascular width measuring algorithm to identify beaded blood vessels and incorporate this retinopathy indicator into the image classification rule.

   At present only 57.2% of the vascular network (by length) is correctly identified. The majority

of the remaining 42.8% are small capillaries. As these capillaries change angle within the retina their retinal appearance becomes non-continuous, and as a result is the primary cause of false positive microaneurysm / haemorrhage classifications. Further investigation into segmentation algorithms is therefore required to segment small non-continuous capillaries.

The presence of new retinal vessels (Neovascularization) that are commonly narrow, tortuous, weak and prone to vascular leakage is an indicator of the retinopathy severity. Identifying new vessels is a challenging task due to their restricted diameter and interweaving appearance. The current vascular segmentation algorithm is incapable of reliably detecting such vessels and would consequently require additional work.

2. **High Resolution Images**

The low resolution of screening images (760 × 570) used within this study may prevent the detection of the earliest clinical diabetic retinopathy indicators. Small vascular variations such as venous beading and microaneurysms may be too subtle to be identified at this scale. If the image resolution is too high however, subtle retinal pigmentation variations may produce unwanted false positive classifications. The effect of the pigment variations can be reduced by filtering or even down-sampling the image to increase contrast, however the effects this may have on early diagnoses is unknown. Further study is therefore required to investigate a range of fundus resolutions, reapplying the presented algorithms at each resolution and comparing the performance to an ophthalmic gold standard evaluation.

3. **Lesion Detection**

This research has concentrated on the detection and classification of the most common clinical lesions indicative of early retinopathy progression (such as microaneurysms, haemorrhages and exudates). A more comprehensive system should also be able to identify lesions or abnormalities associated with different stages of progression, such as cotton wool spots or vitreous haemorrhage (bleeding into vitreous gel caused by new vessel leakage).

9.5 Conclusion

This research has concentrated on the development of techniques to automate retinal analysis and in particular the automated detection of images with referable maculopathy. Several novel algorithms have been presented to address specific research problems within this larger domain, including optic nerve head segmentation, image gradability, lesion segmentation and vascular width measurements. These algorithms when combined with existing methods for foveal location and vascular segmentation have resulted in a reliable integrated system capable of the automated image classification of sight threatening retinopathy (maculopathy).

The role of the system in detecting maculopathy has been evaluated against two applications - 'pre-filtering' and as a 'screening programme'. As a 'screening programme', the system's specificity was 2% under the minimum guidelines proposed by Diabetes UK, and required a full ophthalmic assessed gold standard screening set to confirm results. Therefore, at present, the system is not able to replace human graders and cannot reliably automate patient referrals to ophthalmologists based on retinal images.

As a 'pre-filtering' screening tool capable of excluding images without sight threatening maculopathy from manual grading, the system is fully capable. With a sensitivity of 100%, and specificity of 92%, the system can reduce the number of images requiring human grading by 74.3%.

Bibliography

[1] G.E. Oien and P. Osnes, "Diabetic retinopathy: automatic detection of symptoms from retinal image," *Preceedings NORSIG-95 Norwegian Signal Processing*, 1995.

[2] D.E. Singer, D.M. Nathan, H.A. Fogel, and A.P. Schachat, "Screening for diabetic retinopathy," *Annals of Internal Medicine*, vol. 116, pp. 660–71, 1992.

[3] Imaginis, "History of medical diagnosis and diagnostic imaging," Internet, 08 2004.

[4] World Health Organization, "Diabetes fact," Internet, 08 2004.

[5] R. Klein, B. Klein, S. Moss, M. Davis, and D. Demets, "The wisconsin epidemiologic study of diabetic retinopathy ii. prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years," *Archives of Ophthalmology*, vol. 102, pp. 520–526, 1984.

[6] R. Klein, B. Klein, S. Moss, M. Davis, and D. Demets, "The wisconsin epidemiologic study of diabetic retinopathy iii. prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years," *Archives of Ophthalmology*, vol. 102, pp. 527–532, 1984.

[7] Working party of the British Diabetic Association Report, "Retinal photography screening for diabetic eye disease," Tech. Rep., 2003.

[8] C. Sinthanayothin, J.A. Boyce, T.H. Williamson, H.L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Medicine*, vol. 19, pp. 105–112, 2002.

[9] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 203–210, 2000.

[10] R. Klein and B.E.K. Klein, "Vision disorders in diabetes," *National Institute of Diabetes and Digestive and Kidney Diseases*, pp. 293–338, 1995.

[11] Early Treatment Diabetic Retinopathy Study Research Group, "Photocoagulation for diabetic macular edema etdrs report no. 1," *Ophthalmology*, vol. 103, pp. 1796–806, 1985.

[12] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing (Second Edition)*, Prentice Hall, Upper Saddle River, NJ, 2001.

[13] L.M.T. Collum and D.E. Wallace, *Diabetic retinopathy*, chapter 2.

[14] E.M. Kohner and M. Sleightholm, "Does microaneurysm count reflect severity of early diabetic retinopathy?," *Ophthalmology*, vol. 93(5), pp. 586–589, 1986.

[15] H-C. Chen, *Vascular complications of Diabetes; current issues in pathogenesis and treatment*, chapter 10.

[16] J.J. Kanski, *Clinical Ophthalmology. A systematic approach.*, Butterworth Heinmann, 2003.

[17] icare4u, "Drusen," Internet, 06 2005.

[18] Mayo Clinic staff, "Diabetic retinopathy," Internet, 08 2004.

[19] T. Teng, M. Lefley, and D. Claremont, "Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy," *Medical and Biological Engineering and Computing*, vol. 40, pp. 2–13, 2002.

[20] A. Daxer, "Characterization of the neovascularization process in diabetic retinopathy by means of fractal geometry: diagnostic implications," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 231, pp. 681–686, 1993.

[21] W. Lotmar, A. Freiburghaus, and D. Bracher, "Measurement of vessel tortuosity on fundus photographs," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 211, pp. 49–57, 1979.

[22] P.H. Gregson, Z. Shen, R.C. Scott, and V. Kozousek, "Automated grading of venous beading," *Computers and Biomedical Research*, vol. 28, pp. 291–304, 1995.

[23] H. Littmann, "Zur bestimmung der wahren gro{sse eines objektes auf dem hintergrund des lebenden auges," *Klin Monastbl Augenheilkd*, p. 180:286, 1982.

[24] K. Akita and H. Kuga, "A computer method of understanding ocular fundus images," *Pattern Recognition.*, vol. 15, pp. 431–443, 1982.

[25] C. Sinthanayothin, J.A. Boyce, H.L. Cook, and T.H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Br J. Ophthalmol.*, vol. 83, pp. 902–910, 1999.

[26] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *IEEE Transactions on Medical Imaging*, vol. 23(2), pp. 256–264, 2004.

[27] K. Yogesan, C.J. Barry, et al., "Software for 3-d visualization/analysis of optic-disc images," *IEEE Engineering in Medicine and Biology*, vol. 18, pp. 43–49, 1999.

[28] M. Lalonde, M. Beaulieu, and L. Gagnon, "Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching," *IEEE Trans. Medical Imaging*, vol. 20, pp. 1193–1200, 2001.

[29] Katholieke Universiteit Leuven, "Hough transform," Internet, 08 2004.

[30] D.H. Ballard, M. Marinucci, F. Prioetti Orlandi, A. Rossi-Mori, and L. Tentori, "Automatic analysis of human haemoglobin fingerprints," *Proc., 3rd Meeting, Int'l. Soc. of Haematology, London*, August 1975.

[31] C. Kimme, D. Ballard, and J. Sklansky, "Finding circles by anarray of accumulators," *Communications of the ACM*, vol. 18(2), pp. 120122, February 1975.

[32] S. Lee, *Visual Monitoring of Glaucoma*, Ph.d., Robotics Research Group Department of Engineering Science, University of Oxford, 1991, Available on micro-fiche.

[33] B. Ege, O. Larsen, and O. Hejlesen, "Detection of abnormalities in retinal images using digital image analysis," *In Proceedings of the 11 th Scandinavian Conference on Image*, pp. 833–840, 1999.

[34] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes. active contour models," *Int. J. Comput. Vision*, vol. 1, pp. 321–331, 1987.

[35] D. Terzopoulos and D. Metaxas, "Dynamic 3d models with local and global deformations: deformable superquadrics," *IEEE Trans. PAMI*, vol. 13, pp. 703–714, 1991.

[36] A.K. Jain, Y. Zhong, and M. Dubuisson-Jolly, "Deformable template models: A review," *Signal Processing*, vol. 71, pp. 109–129, 1998.

[37] F.A. Velasco and J.L. Marroquin, "Robust parametric active contours: the sandwich snakes," *Machine Vision and Applications*, vol. 12, pp. 238–242, 2001.

[38] H. Eviatar and R.L. Somorjai, "A fast, simple active contour algorithm for biomedical images," *Pattern Recognition Letters*, pp. 969–974, 1996.

[39] F. Mendels, C. Heneghan, P.D. Harper, R.B. Reilly, and J-Ph. Thiran, "Extraction of the optic disc boundary in digital fundus images," in *Proc. 1st Joint BMES/EMBS Conf.*, 1999, p. 1139.

[40] Y-L. Hu, W.J. Rogers, D.A. Coast, C.M. Kramer, and N. Reichek, "Vessel boundary extraction based on a global and local deformable physical model with variable stiffness," *Magnetic Resonance Imaging*, vol. 16, pp. 943–951, 1998.

[41] A. Blake and M. Isard, *Active Contours*, Springer, 1998.

[42] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C*, 1992.

[43] C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

[44] M. Lalonde and L. Gagnon, "Variable neighbourhood search for geometrically deformable templates," in *Proc. Int. Conf. Patt. Rec.*, Québec, Canada, 2002, pp. 689–692.

[45] A. Hunter, J. Lowell, D. Steel, A. Basu, and R. Ryder, "Non-linear filtering for vascular segmentation and detection of venous beading," Internal Report 08/02, Department of Computer Science, University of Durham, Science Labs, South Rd., Co. Durham, DH1 3LE, UK., 2002.

[46] L. Pedersen, M. Grunkin, B. Ersboll, K. Madsen, et al., "Quantitative measurement of changes in retinal vessel diameter in ocular fundus images," *Pattern Recognition Letters*, vol. 21, pp. 1215–1223, 2000.

[47] Early treatment diabetic retinopathy study research group, "Fundus photographic risk factors for progression of diabetic retinopathy," *Ophthalmology*, vol. 98, pp. 823–833, 1991.

[48] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating the blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 203–210, 2000.

[49] AMP. Hamilton, M.W. Ubig, and P. Polkinghome, *In Management of Diabetic retinopathy*, chapter 3, p. 123, BMJ publishing group, 1996.

[50] M.H. Goldbaum, N.P. Katz, M.R. Nelson, and L.R. Haff, "The discrimination of similarly colored objects in computer images of the ocular fundus," *Investigative Ophthalmology and Visual Science*, vol. 31(4), pp. 617–623, 1990.

[51] N.P. Ward, S. Tomlinson, and C.J. Taylor, "Image analysis of fundus photographs. the detection and measurement of exudates associated with diabetic retinopathy," *Ophthalmology*, vol. 96(1), pp. 80–86, 1989.

[52] H. Wang, W. Hsu, K. Guan, and M. Lee, "An effective approach to detect lesions in color retinal images," *Proc. IEEE Conf. on Computer Vision an Pattern Recognition*, vol. 2, pp. 181–6, 2000.

[53] A. Osarch, M. Mirmehdi, B. Thomas, and R. Markham, "Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural network," in *Medical Image Understanding and Analysis*, E. Claridge and J. Bamber, Eds. 2001, pp. 49–52, BMVA Press.

[54] Y.W. Lim and S.U. Lee, "On the colour image segmentation algorithm based on the thresholding and the fuzzy c-means clustering techniques," *Pattern Recognition*, vol. 23, pp. 1990, 1990.

[55] A. Frame, P.E. Undrill, M. Cree, J. Olson, K. McHardy, P. Sharp, and J.V. Forrester, "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms," *Computers in Biology and Medicine*, vol. 28, pp. 225–238, 1998.

[56] T. Spencer, J.A. Olson, K.C. McHardy, P.F. Sharp, and J.V. Forrester, "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescien angiograms of the ocular fundus," *Computers and biomedical research*, vol. 29, pp. 284–302, 1996.

[57] R.J. Henery, "Classification," pp. 6–16, 1994.

[58] D. Michie, D.J. Spiegelhalter, and C.C. Taylor, "Introduction," pp. 1–5, 1994.

[59] J.M.O. Mitchell, "Classical statistical methods," pp. 17–28, 1994.

[60] B. Ege, O. Larsen, O. Hejlesen, et al., "Screening for diabetic retinopathy using computer based image analysis and statistical classification," *Computer methods and programs in biomedicine*, pp. 165–175, 2000.

[61] Computing Research Association, "K nearest neighbor," Internet, 09 2004.

[62] G. Gardner, D. Keating, T.H. Williamson, and A.T. Elliot, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool," *British Journal of Ophthalmology*, vol. 80, pp. 940–944, 1996.

[63] A. Hunter, J. Lowell, J. Owens, and L. Kennedy, "Quantification of diabetic retinopathy using neural network and sensitivity analysis," in *Proceedings of Artificial Neural Networks in Medicine and Biology*, 2000, pp. 81–86.

[64] J. Lowell, A. Basu, A. Hunter, D. Steel, L. Kennedy, and R. Ryder, "A neural network system for diagnosis of diabetic retinopathy by white lesion detection and classification," *Diabetes UK*, 2002.

[65] P.H. Scanlon, R. Malhotra, G. Thomas, C. Foy, J.N. Kirkpatrick, N. Lewis-Barned, B. Harney, and S.J. Aldington, "The effectiveness of screening for diabetic retinopathy by digital imaging photography and technician ophthalmoscopy," *Diabetes UK. Diabetic Medicine*, vol. 20, pp. 467 474, 2003.

[66] NHS, "A model for a photographic diabetic retinopathy screening programme," Internet, 10 2004.

[67] D.B. Usher, M. Himaga, M.J. Dumskyj, J.F. Boyce, A. Sabate-Cequier, and Williamson T.H., "Automated assessment of digital fundus image quality using detected vessel area," *Procedings of medical image understanding and analysis the university of sheffield*, pp. 81–84, 2003.

[68] M. Himaga, D. Usher, and J.F. Boyce, "Retinal blood vessel extraction by using multi-resolution matched filtering and directional region growing segmentation," *Proc. IAPR Workshop on Machine Vision Applications, Nara, Japan*, pp. 244–247, 2002.

[69] S.C. Lee and Y. Wang, "Automatic retinal image quality assesment and enhancements," in *Proc SPIE Vol 3661 Med Imaging:Img Processing*, 1999, pp. 1581–90.

[70] M. Lalonde, M. Beaulieu, and L. Gagnon, "Automated visual quality assessment in optical fundus images," in *Vision Interface*, 2001, pp. 259–264.

[71] S.M.B. Rassam, V. Patel, 0. Brinchmann-Hansen, O. Engvold, and E.M. Kohner, "Accurate vessel width measurement from fundus photographs: a new concept," *British Journal of Ophthalmology*, vol. 78, pp. 24–29, 1994.

[72] A.R. Sharrett, L.D. Hubbard, L.S. Cooper, P.D. Sorlie, et al., "Retinal arteriolar diameters and elevated blood pressure: the atherosclerosis risk in communities study," *American Journal Epidemiology*, vol. 150, pp. 263–270, 1999.

[73] T.Y. Wong, R. Klein, B.E. Klein, J.M. Tielsch, et al., "Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease and mortality," *Survey of Ophthalmology*, vol. 46, no. 59-80, 2001.

[74] N. Chapman, G. Dell'omo, M.S. Sartini, and N. Witt, "Peripheral vascular disease is associated with abnormal arteriolar diameter relationships at bifurcations in the human retina," *Clinical Science (London)*, vol. 103, pp. 111–116, 2002.

[75] T.Y. Wong, R. Klein, D.J. Couper, L.S. Cooper, et al., "Retinal microvascular abnormalities and incident stroke: the atherosclerosis risk in communities study," *Lancet*, vol. 358, pp. 1134–1140, 2001.

[76] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 263–369, 1989.

[77] O. Brinchmann-Hansen and H. Heier, "The apparent and true width of the blood column," *Acta Ophthalmologica, Supplement*, vol. 179, pp. 29–32, 1986.

[78] H. Neame, "A method of estimating the calibre of retinal arteries in the living eye by means of opthalmoscope, illustrated results in some normal and pathological cases," *Transactions of the ophthalmological societies of the united kingdom*, vol. 56, pp. 155–162, 1936.

[79] T. Behrendt, "Scanning densitometer for photographic fundus measurements," *American Journal of Ophthalmology*, vol. 62, pp. 689–693, 1966.

[80] A. Kagan, E. Aurell, and G. Tibblin, "Signs in the fundus oculi and arterial hypertension. unconventional assessment and significance," *Bull WHO*, vol. 36, pp. 231–241, 1967.

[81] J.V. Hodge, J.C. Parr, and G.F.S. Spears, "Comparison of methods of measuring vessel widths on retinal photographs and the effect of fluorescein injection on apparent retinal vessel calibre," *American Journal of Ophthalmology*, vol. 68, pp. 1060–1068, 1969.

[82] O. Brinchmann-Hansen and O. Engvold, "Microphotometry of the blood column and light streak on retinal vessels in fundus photographs," *Acta Ophthalmologica, Supplement*, vol. 179, pp. 9–19, 1986.

[83] O. Brinchmann-Hansen and H. Heier, "Theoretical relationships between light streak characteristics and optical properties of retinal vessels," *Acta Ophthalmologica, Supplement*, vol. 179, pp. 33–37, 1986.

[84] P.H. Gregson, Z. Shen, R.C. Scott, and V. Kozousek, "Automated grading of venous beading," *Computers and Biomedical Research*, vol. 28, pp. 291–304, 2000.

[85] N. Chapman, N. Witt, X. Goa, A. Bharath, A.V. Stanton, S.A. Thom, and A.D. Hughes, "Computer algorithms for the automated measurements of retinal arteriolar diameters," *British Journal of Ophthalmology*, vol. 85, pp. 75–79, 2001.

[86] L. Zhou, M.S. Rzeszotarski, L.J. Singerman, and J.M. Chokreff', "The detection and quantification of retinopathy using digital angiograms," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, December 1994.

[87] X.W. Goa, A. Bharath, A. Stanton, A. Hughes, N. Chapman, and S. Thom, "Quantification and characterisation of artieries in retinal images," *Computer Methods and Programs in Biomedecine*, vol. 63, pp. 133–146, 2000.

[88] L. Gang, O. Chutatape, and S.M. Krishnan, "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order gaussian filter," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, February 2002.

[89] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and L. Kennedy, "Measurement of retinal vessel widths from fundus images based on 2-d modeling," *IEEE Transactions on Medical Imaging*, vol. 23(10), pp. 1196– 1204, 2004.

[90] D. Usher, M. Dumskyjs, M. Himaga, T.H. Williamson, S. Nussey, and J. Boyce, "Automated detection of diabetic retinopathy in retinal images: a tool for diabetic retinopathy screening," *Diabetic Medcine*, vol. 21, pp. 84–90, 2003.