Vessel identification in diabetic retinopathy

Thomas Bart Teng

A thesis submitted in partial fulfilment of the requirements
of Bournemouth University for the degree
of Doctor of Philosophy

2003

Bournemouth University

Vessel identification in diabetic retinopathy

Abstract

Vessel identification in diabetic retinopathy

Thomas Bart Teng

Bournemouth University

Diabetic retinopathy is the single largest cause of sight loss and blindness in 18 to 65 year olds. Screening programs for the estimated one to six percent of the diabetic population have been demonstrated to be cost and sight saving, however there are insufficient screening resources. Automatic screening systems may help solve this resource short fall. This thesis reports on research into an aspect of automatic grading of diabetic retinopathy; namely the identification of the retinal blood vessels in fundus photographs. It develops two vessels segmentation strategies and assess their accuracies.

A literature review of retinal vascular segmentation found few results, and indicated a need for further development. The two methods for vessel segmentation were investigated in this thesis are based on mathematical morphology and neural networks. Both methodologies are verified on independently labeled data from two institutions and results are presented that characterise the trade off between the ability to identify vessel and non-vessels data.

These results are based on thirty five images with their retinal vessels labeled. Of these images over half had significant pathology and or image acquisition artifacts. The morphological segmentation used ten images from one dataset for development. The remaining images of this dataset and the

entire set of 20 images from the second dataset were then used to prospectively verify generaliastion. For the neural approach, the images were pooled and 26 randomly chosen images were used in training whilst 9 were reserved for prospective validation.

Assuming equal importance, or cost, for vessel and non-vessel classifications, the following results were obtained; using mathematical morphology 84% correct classification of vascular and non-vascular pixels was obtained in the first dataset. This increased to 89% correct for the second dataset. Using the pooled data the neural approach achieved 88% correct identification accuracy. The spread of accuracies observed varied. It was highest in the small initial dataset with 16 and 10 percent standard deviation in vascular and non-vascular cases respectively. The lowest variability was observed in the neural classification, with a standard deviation of 5% for both accuracies.

The less tangible outcomes of the research raises the issue of the selection and subsequent distribution of the patterns for neural network training. Unfortunately this indication would require further labeling of precisely those cases that were felt to be the most difficult. I.e. the small vessels and border conditions between pathology and the retina.

The more concrete, evidence based conclusions, characterise both the neural and the morphological methods over a range of operating points. Many of these operating points are comparable to the few results presented in the literature. The advantage of the author's approach lies in the neural method's consistent as well as accurate vascular classification.

Contents

This page intentionally contains only this sentence.

List of Figures

This page intentionally contains only this sentence.

List of Tables

This page intentionally contains only this sentence.

Dedication and Acknowledgements

I would like to thank my supervisors, Prof. Denzil Claremont and Dr. Martin Lefley, for their advice, dedication and patience. I salute my family for their love, and the Free Software Foundation GNU Public License for all the quality tools it has provided.

This page intentionally contains only this sentence.

List of Abbreviations

AH   Adam Hoover, image labeller from Hoover et al. (2000).

ANN   Artificial Neural Network.

CSME   Clinically Significant Macular Edema.

DR   Diabetic Retinopathy.

HA   HAemorrhages.

IRMA   Intra Retinal Micro-vascular Abnormalities.

MO   Macular Oedema.

NPDR   NonProliferative Diabetic Retinopathy.

PDR   Proliferative Diabetic Retinopathy.

RBH   Royal Bournemouth Hospital

ROC   Receiver Operating Characteristic.

TPF   True Positive Fraction or sensitivity.

TNF   True Negative Fraction or specificity.

This page intentionally contains only this sentence.

Chapter 1

Introduction

1.1 Summary of research and thesis outline

This thesis explores the identification of blood-vessels in images of the photosensitive tissues of the human eye. Artificial neural networks and mathematical morphology are investigated to classify picture-elements of digital retinal-images into two classes; vessel and non-vessel. Reliable identification of the vessels is an important goal in its own right and forms a significant step in the processing of retinal images for the analysis of injury or for the characterisation of images in screening.

The retinal blood-vessels are susceptible to the abnormal metabolism of glucose. Over time, this metabolic malfunction known as diabetes alters the structure, and consequently the function, of retinal vessels. These changes to the vessels trigger a sequence of injuries or lesions to the retina itself that are collectively known as Diabetic Retinopathy (DR). The diagnosis, monitoring and screening of diabetic injuries to the retina overwhelms current ocular health care, causing loss of sight and blindness. The author believes that the identification of the blood vessels constitutes an important step or building block in the design of systems for automated

screening or analysis.

A review of the published literature reveals that there is no reference method for the segmentation of the retinal vessels. It also reveals there are many segmentation strategies but little evidence to discern between their performance and reliability. This thesis provides an objective characterisation, or a transportable description of the performance of mathematical morphology and neural network based vascular segmentation.

1.1.1   Structure of thesis

This thesis follows a conventional structure; the remainder of this chapter will thus introduce diabetes, its effects on the retina and the implications for ocular health-care. The following chapter (Chapter 2) will collate, analyse and condense previous research for the identification and analysis of retinal vessel data. After this review, the morphological and neural network methods for segmentation of the retinal-vessels developed by the author are documented (Chapters 3 and 4). Each of these chapters will also describe the accuracies of applying these methods to the retinal images available in this research. Chapter 5 provides a summary of the segmentation methods and compares morphological and neural-based segmentation. The final chapter also draws conclusions which are used to highlight further work in an effort to answer some of the questions raised by this thesis.

1.2   Diabetes and screening

Diabetes is hallmarked by the dysfunctional metabolism of glucose and arises out of the ineffective production or action of the glucose regulating enzyme called insulin. The distinction between action and production of insulin is reflected in the two types of diabetes, type I corresponding to ineffective production, and type II corresponding to the ineffective action of

insulin. Both of these types of diabetics benefit from screening. The use of the word screening in this respect means the mass testing of asymptomatic diabetics, and not the non-diabetic population. Ocular screening is performed to detect, treat and prevent the harmful consequences that chronic diabetes has on sight. This vision loss arises from vascular changes in the retina which in turn damages the other retinal tissues.

The field of Diabetic Retinopathy screening is complex due to the interplay between the natural history of DR, the modalities used to detect it, their costs and the shifting demographic and its effects upon the epidemiology of diabetes. Despite these variable conditions every diabetic person is at risk from developing retinal complications. These complications often become visible on the fundus; the visual aspect of the back of the eye dominated by retinal tissues, see Figure 1.1 . A regular examination of the fundus is necessary to monitor or establish the type and extent of the complications. Since early diagnosis combined with treatment have been demonstrated to be highly beneficial in medical as well as financial analyses, screening for DR has become more important (WHO, 1989). Before details of the diagnostic aspects of DR are given the following section will discuss the number of people affected by diabetes and how the DR they develop depends on the type of diabetes they have.

Every human is at risk of developing diabetes and the chances of being diagnosed with the condition increase with age. Attempts at measuring how much diabetes is present in the population (prevalence) result in estimates between 1 to 6%. Approximately half of these people are undiagnosed and thus outside of the health care system (Harris et al., 1987).

The information on the diabetic status of the population describes the type and extent of diabetic lesions expected in the retina. Information on the distribution and likely-impact of lesions helps the design of systems processing that data. The epidemiology of DR is well described by two population based studies: the Early Treatment of Diabetic Retinopa-
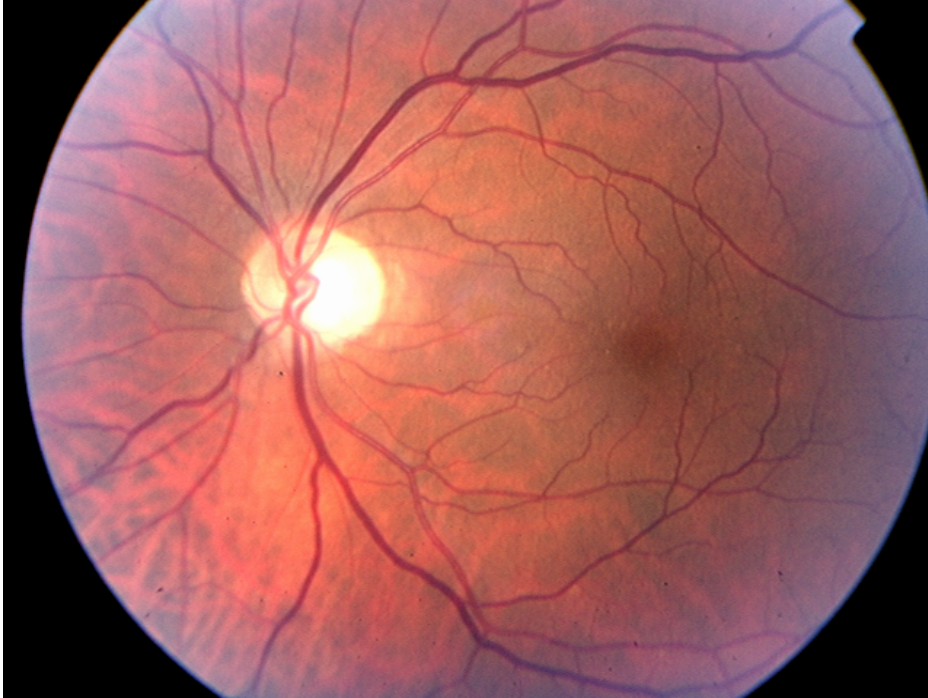
Figure 1.1: Reflectance fundus image of a healthy retina. Large, bright and circular feature left of centre image is the optic disc.

thy Study (ETDRS) and the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR). These two studies used the most sensitive and repeatable diagnostic-protocol available and provide the evidence upon which todays treatment and screening procedures are based. These findings shall be summarised before a description of the grading methods / diagnostic protocols is given.

1.2.1   Prevalence of diabetes

In a random sample of the North American population, diabetes was detected in 6.6% of those screened (Harris et al., 1987), in half of these diabetics it was the first diagnosis of their condition. The other half had a known history of diabetes, one quarter of these were using insulin. The

use of insulin is associated with type I diabetes, but is not a determinant thereof. The prevalence of diabetes also varies with age; between the ages of 20 years and 44 years the prevalence is lowest at 2%. In the 55 to 74 years-old range diabetics account for approximately 15% of the population. The prevalence in the older age group is expected to rise due to: increased incidence, shifting demographics and reduced morbidity of diabetes.

Incidence rates of diabetes have indeed increased (Ruwaard et al., 1996). This study monitored 1% of the population of The Netherlands and found 1.5 persons per 1000 per year were newly diagnosed with diabetes. This represented a significant increase over the previous decade by about a third. Also noted in the study is the inability to distinguish changes in aetiology (or causes) and detection methods as the cause of the increase.

1.2.2   Financial impact of sight loss

For diabetics in health care; best practice and current guidelines for diabetic-ocular screening are based on, and justified by, the benefits of timely treatment. These recommendations suggest diabetic persons have an annual ophthalmic examination (American Diabetes Association, 1999; Aiello et al., 1998b). There is an exception to this guideline for newly diagnosed type I diabetics; they do not need ocular screening after the first examination for three to five years. In general however, approximately one to six percent of the population would require annual screening. The demand this screening exerts on health-care overwhelms its capacity to supply (WHO, 1989). Progression rates of DR with and without the positive effects of treatment demonstrate that screening is not only cost effective; i.e. is perceived to be worthy of funding, but also that it is cost saving in that it actively prevents expenditure. From a governmental point of view screening would unlock funds that are already being spent.

These financial models incorporate prevalences outlined in the previ-

ous section. They also incorporate clinical assumptions of disease progression, treatment and its outcomes that are detailed in section 1.5 (page 20). Models developed for the United-States agree that screening for type I diabetics is cost saving, however they vary in their financial forecasts for type II diabetics. Predictions of income generation for type I diabetes such as US$100 Million 1994 per anum (Javitt et al., 1994) or US$1100 Million 1991 incremental-savings per decade estimated by Dasbach et al. (1991) enhance the medical and social justifications. Both Javitt et al. (1994) and Dasbach et al. (1991) predict significant cost savings for resource-intensive annual photographic-screening through dilated pupils. The disagreement between the models arises out of assumptions for successful screening of type II diabetics with macular oedema.

Macular oedema; or the accumulation of fluid in the intraretinal space, is an important lesion type in DR. Its impact is illustrated by the swing from net income generation to an outright expense depending on its inclusion when screening type II diabetics. During clinical trials for the early treatment of macular oedema, treatment outcomes were considered so beneficial that it was un-ethical to continue and finish the non-treatment arm of the trial (ETDRS Research Group, 1991a) in its presence.

Javitt et al. (1994) includes Macular Oedema (MO) as a valid endpoint for screening, consequently it is estimated that treating type II patients would generate savings of US$250 Million 1994 per year. In contrast when MO is not deemed a valid target for screening (Dasbach et al., 1991) approximately US$274 Million 1991 is lost every ten years for this cohort. In terms of years of sight saved, there is a fourfold increase in the number of years of sight saved, from 48,500 years to approximately 200,000 years if MO is screened.

1.2.3 Fundus screening and ocular health care for diabetics

Current preferred practice is based on in-vivo examination of the ocular fundus (American Diabetes Association, 1999). Supplemental to live-examination is stereoscopic colour fundus photography. These images, recorded on photographic slides or negatives, are recommended when a permanent record is necessary or useful. For example: in planing therapy, in borderline cases and during active disease proliferation. Together with the examinations of the fundus, other clinical tests to assess, visual acuity, blood metabolites and the other major tissues of the eye are also recommended to form part of a screening episode.

The screening advice given by Diabetes Associations and Ophthalmology Institutions varies throughout the developed world. American health organisations such as the Diabetic Association and Academy of Ophthalmology recommend resource and man-hour intensive programs (American Academy Ophthalmology Prefered Practice Patterns Committee Retina Panel, 1998; American Diabetes Association, 1999). Both suggest annual screening using photography via dilated pupils (midriasis). In England committees have also concluded that an annual examination is necessary. No stipulations on the imaging modality have been set (National Screening Committee, 2000; National Institute Clinical Excelence, 2002), but they have recommended digital photography to be phased in. Unfortunately, from a automated screening perspective, a medical history is also stipulated by the recommendations.

1.3 Introduction to Diabetic Retinopathy

The remainder of this chapter will describe many of the anatomical structures that are often the subject of automated image analysis. It will also

describe the manifestations of diabetic retinopathy as well as their detection and grading. Details are given on the cellular changes in the vessels caused by the abnormal chemical environment of diabetes, and how this leads to changes visible to the instruments of the ophthalmologist and retinal specialist. The energy and transducers used to detect these changes in vascular morphology and the lesions to the retinal tissues determine clinical (the physical measurements), medical (patient management) and diagnostic (data interpretation) practices. In the study of the author, it is white light in the visible spectrum that is detected and recorded by silver-halide film or charge coupled devices. Note this shapes the perspective of this research yet DR may also be investigated via images derived spectroscopically (Beach et al., 1999), from scanning laser ophthalmoscopy, electro retinograms, confocal imaging, angiographic techniques and stereophotography (Kiri et al., 1996). The last sections of this chapter will describe how the progression of diabetic retinal pathology is diagnosed and measured.

1.4   Anatomy and Physiology of the retina

The retina lies between the central, clear vitreous body and the choroid, see Figure 1.2. The vitreous body is filled with a translucent gel whose composition varies concentrically; its most liquid phase being central which gradually turns into a fibrous network as it nears the retina. This limiting aspect of the retina is formed by the extension of foot like processes from the matrix of structural cells supporting vascular and neuro-sensory aspects of the retina proper. The most exterior retinal stratum, going from the centre of the eyeball outwards, is the Retinal Pigment Epithelium or RPE. This is one of the blood retinal barriers which involved in the homeostasis of the retina proper. The lesions of early DR, or mild to moderate
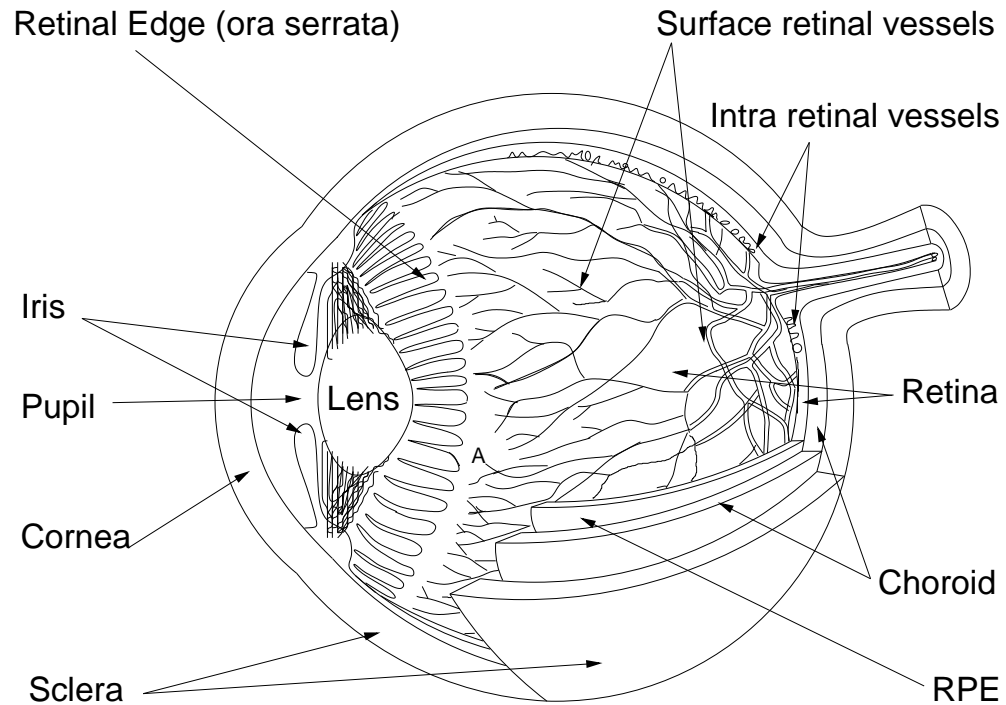
Figure 1.2: Schematic illustration of the eye ball and the organisation of the tissues within it.

cases, are mostly found between the structural cells near the vitreous, limiting the retinal interiorly, and the external RPE.

Like all tissues in the body the retina has characteristics such as size, colour, shape and distribution corresponding to its functional divisions. The retina proper has a layered structure, see illustration in Figure 1.3. The retinal vessels may be also characterised by the number of branches they have undergone. The central retinal vein and central retinal artery are shown within the bundle protruding from the top right of the eyeball in Figure 1.2. The central retinal vein and artery are not true arteries or veins; they bear closer resemblance to arterioles and venules in the proportions and makeup of their vessel wall tissues. The vessels that enter the retina are branches of the ophthalmic artery or vein which in turn branches

from the internal carotid vascular subsystem. The central retinal vein and artery are of the smaller primary branches of the ophthalmic vessels and travel to the retina by penetrating the optic nerve bundle, within which they lie centrally until reaching the retina where they start branching.

These two vessels supply and drain the retina, however the retina's metabolic and catabolic needs are also dependent on the choroidal circulation of the uveal tract. The uveal tract is supplied by branches of the ophthalmic artery, such as the long or short ciliary vessels, and is drained by vorticose veins which (like the cilliary vessels) puncture the scleral tissues of the posterior hemisphere. The short ciliary vessels perforate the sclera, around the junction of the optic bundle to the eye, into the uveal tract. The uveal tract extends anteriorly towards the iris via the ciliary body, which is also involved in the homeostasis of the retina via its blood-retina barrier. When the RPE is lightly pigmented the vessels of the uveal tract in the choroid often become visible.

Retinal function is thus dependent on the simultaneous and normal perfusion of both the uveal and the retinal circulation. There is an exception to this, the fovea or area of highest visual acuity is primarily dependent on the choroidal circulation only as there are no retinal vessels in this area. The fovea is located temporal to the optic disk and can be seen on the vertical midline and just below the horizontal midline in Figure 1.4. The fovea lies at the centre of the macula lutea, a region pigmented by xanthophyll, this is also visible surrounding the fovea in Figure 1.4.

The first dichotomous division of the central retinal arteries and veins occurs on entry to the retina at the optic disk or blind spot. The vessels service four distinct quadrants; the inferior and superior, nasal or temporal. These four quadrants are separately perfused. The vessels within these quadrants continue dividing until their diameter reaches the 8μm characteristic of a capillary. Most of the divisions of the central retinal vein and artery are visible on the images which include the optic disk (Figures 1.1

Vitreous

Vascular lamina

Retino capillaris

Retina

Avascular Retina

RPE

Choidocapillaris

Choroid

Vascular lamina

Vessel Type

Venous

Arterial

Sclera
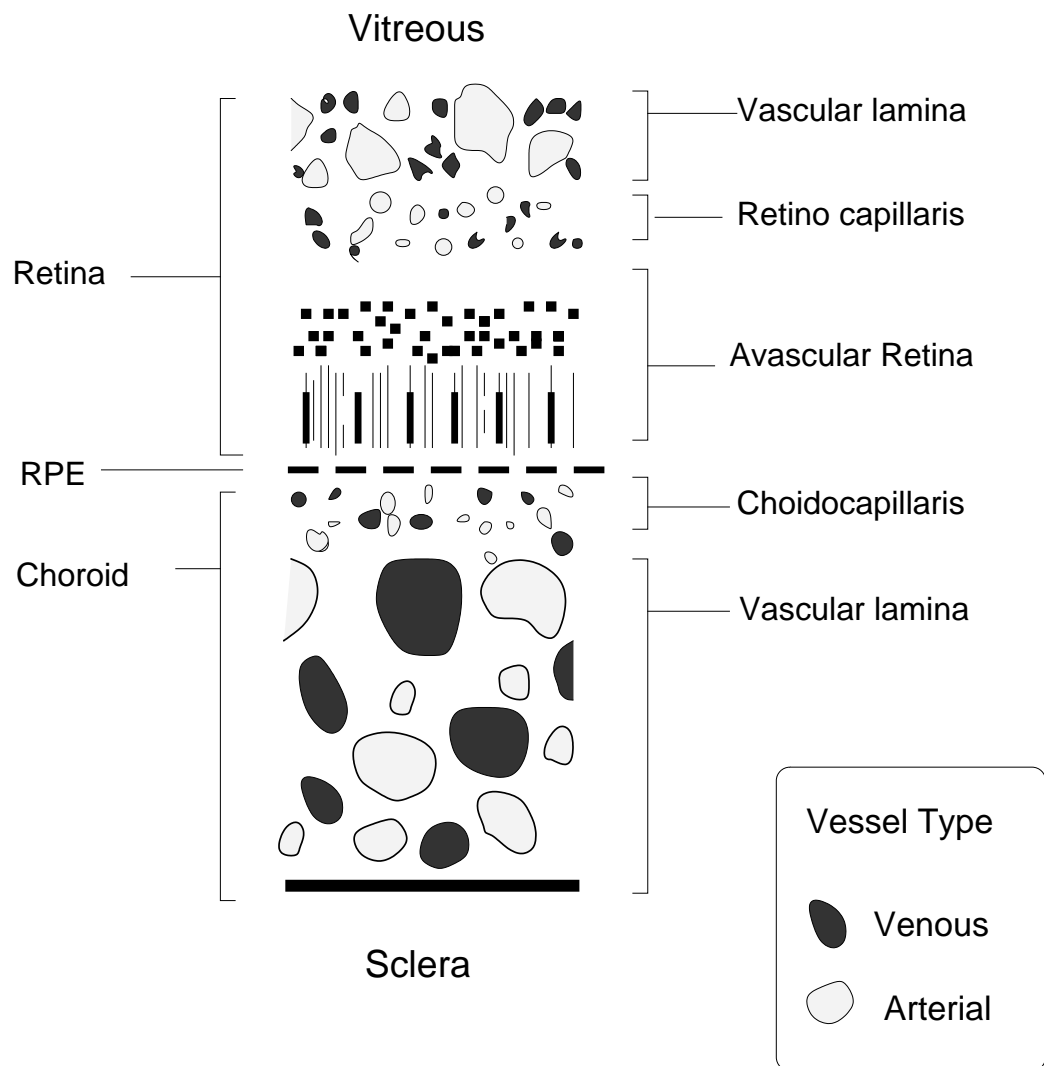
Figure 1.3: Schematic illustration of a retinal crossection showing the avascular neuro-sensory retina between the choroidal and retinal vessel-networks.
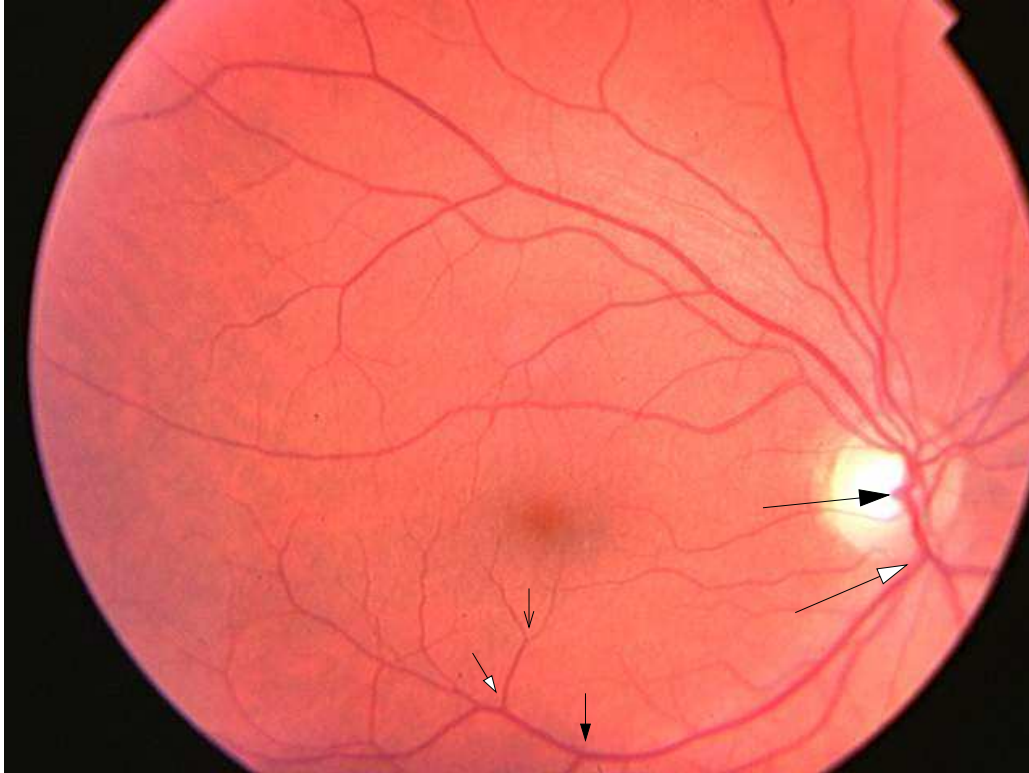
Figure 1.4: Fundus photograph with bright optic disk at right.Visible too, are the first (large black arrow) second (large white arrow) third (small black arrow), fourth (small white) and fifth (small open) divisions of the inferior lateral vein

and 1.4). Note that the veins are darker and of larger calibre. All retinal vessels lie in the path of light reaching the photosensitive aspect of the retina, and thus cast a non-perceptible shadow on the retina.

1.4.1   Colour and pigments of the retina

Retinae are not uniformly pigmented, patches of irregular intensity, area and shape are subject to variability even within individual fundus images; compare Figures 1.1 and 1.4. The substance of greatest light absorbance

coefficient is melanin. Melanin is present in the retinal pigment epithelium and the choriod, its absorbance decreases uniformly over the visual range, 400-700nm. Macular pigment, which is distinct from the phototransductive pigments in the rod and cone cell processes, does not interact with light of wavelengths greater than 530nm (Kilbride et al., 1989).

Oxyhaemoglobin and deoxyhaemoglobin, the greatest light absorbers of the retinal vessels, have absorbance spectra that peak in the 500-600nm range (Flewelling, 1995). The perceptual difference in colour between venous and arterial blood is attributable to the steeper decline in absorbance of oxyhaemoglobin after 600nm, and thereby more effectively reflecting the red component of incident white light. As figure 1.5 shows the non-pigmented vessel walls (of the larger veins and arteries) have a minor but often distinguishable absorbance from the general fundus (Patel, 1995).

The pigment of the RPE and choriod is visible at the macula. This area is also variably pigmented in itself, both its translucency and its own 'yellowish' colouration are perceptible as the darkened area just below the centre of Figure 1.4. In the centre of the macula is the fovea, and is the area of highest visual acuity. The fovea is avascular and additionally pigmented. Moving from the outer limit of the macula towards the centre of the fovea the interior layers of the retina are progressively retracted, so that at the fovea only the rod cell and cone cell processes and the associated axons lie between the RPE and the clear vitreous of the posterior chamber. Here at the fovea the retina is devoid of vessels, and its metabolic needs are met by diffusion and active transport through the RPE only. The sensory stratum of the rest of the retina has a dual source of perfusion, as both the capillaries within the interior two thirds of the retina (Adler, 1992) and the choirocappilaris are involved. This vasculated volume is approximately 0.4mm thick throughout most of the retina.

1.4.2   Anatomy of the larger retinal blood vessels

The retinal macrovasculature is tubular, the cross-sectional profile is approximately round or elliptical and irregular. The central-retinal artery is about 280μm in diameter (Saude, 1993), and divides to give rise to the first true retinal arteries. After the first division, vessels lie on the optic disk and are about one twelfth its size (1600μm). As most images taken of the retina are not length calibrated this may serve as an approximation. Their distribution or arbor over the fundus has been described empirically by Toussaint and Danis (1970) and in terms of fractal geometry (Landini et al., 1995; Daxer, 1993a; Mainster, 1990).

1.4.2.1   Structure of the larger vessels

The anatomy of an individual retinal-vessel-segment is dimensioned according to its branching level with the retinal veins being larger in calibre than their arterial counterparts. The larger or non-capilary vessels of the retina, are better termed arterioles or venules (Adler, 1992; Patel, 1995) in order to reflect their function as resistance vessels. However, convention has established that the terms 'vein' and 'artery' used to address these vessels, and this terminology will be used in the text.

   The retinal arteries, veins and micro-vasculature are composed of concentric layers wrapping the blood column. These structures have been surmised from the descriptions of the central retinal vessels in Ikui et al. (1964). Innermost and adjacent to the blood column are the endothelial membrane cells which rest upon an elastic membrane. Exterior to these inner layers is a smooth muscle layer which in turn is wrapped by a fibrous tunic. The central retinal vein and artery differ in the thickness of the two innermost layers, both the endothelial and smooth muscle layers being thinner in veins.

   The arteries and veins of the retina are capable of altering in tone and

diameter. The retinal vessels change diameter over and above the pulsatile changes induced by each heart beat (Chen et al., 1994; Dumskyj et al., 1996). These slower changes are attributed to the rhythmical variation in smooth muscle tone. This vasomotion is poorly understood and has only been measured once (Chen et al., 1994), however it has been known of and observed since Michaelson (1948). The healthy retinal vasculature also changes diameter due to posture (Hauge and Hill, 1988), altitude (Meehan et al., 1990) and hyperoxia (Delori, 1988).

In 1994, Chen et al used electrocardiographic gating to demonstrate 4.8 and 3.2 percent changes in venous and arterial vessel diameter over the cardiac cycle. By using the images from the same phase of the cardiac cycle, vasomotive differences in diameter fluctuated 3.7 and 2.6 percent of arteries and veins respectively. Although not made explicit in Chen et al's report, it is clear that pulsatile changes interact with the vasomotive dynamic. Further study is required to elucidate the relationship (if any) between vasomotive and the cardiac cycles.

Postural changes elicit a discordant calibre changes. This change was characterised as autoregulatory and non-autoregulatory or passive. Calibres distend or contract with increases and decreases in pressure associated with supine and seated postures (Hauge and Hill, 1988). The change was bimodal, i.e. there were two distinct reactions, diameter changes go either way. This bi-directionality of response was significant because it occurred randomly within the same individual, and was not attributable to measuring error. The duration of the response was not discussed, neither was the clear involvement of vasomotion.

Induced hyperoxia, by breathing 100% oxygen, constricts veins and arteries by 14 and 10 percent of their diameter respectively (Delori, 1988). This constriction was obtained for a single subject, and was persistent for the duration of the hyperoxia. Two notable results stand out. Firstly there was no increase in oxygen saturation of the arterial blood during hyper-

oxia. Secondly these detailed scans of vessel diameter prior to oxygen saturation measurement failed to establish baseline vascular calibres, again possibly due to vasomotive involvement.

1.4.2.2 Distribution and arbour of the retinal macro-vasculature

On visual inspection the organisation or higher level structure of the retinal vessel network appears to mimic some of the fractal motifs found in nature. The patterns of branching plants, diffusion limited aggregation and dentritic drainage are evocative of the retinal vascular arbour. Some general observations on the retinal macro-vasculature, unsubstantiated by empirical evidence, are now presented. Vessels divide into two and have a general radial orientation with respect to the optic disk. Tapering or smooth reduction in vessel calibre is not perceptible between segments, a segment being defined as a vessel tract between two ramii. Saude (1993) describes a tendency for retinal veins to lie interior relative to arteries, whilst Berry et al (1995) have retinal arteries lying superficial to veins at crossings, in concordance with Yamamoto and Yokouchi (1976) 's characteristics-of-crossing-phenomena. However, vessels may even twist about round each other, especially on or near the optic disk, see Figures 1.6.

The retinal vascular network is part of a greater distribution network. Distribution networks are common to most multicellular life and are reflective of, and structurally intrinsic to, the function and organisation of an organism. Starting with some basic assumptions common to all biological transport networks West et al. (1997) derived scaling laws applicable not only to the vasculature from which they were obtained, but also to many biological and ecological variables. Specifically they demonstrate how biological transport systems themselves scale. Implicated are two ratios of interest to retinal vascular disease. These ratios, before and after a vessel branch, of vessel calibres and segment-lengths, are theorised to have

a fixed value, and thus self similar or fractal. The theory is supported by observations in independent studies that have found this scale invariance in measurements of the fractal dimension of the retinal vascular system (Landini et al., 1995; Daxer, 1993a). Although the actual value of the ratio was not given, selfsimilarity would suggest its existence.

1.4.2.3   Structure of the retinal micro-vasculature

The capillaries have a similar layered structure to the arteries and veins. However, the pulsatile, vasomotive or autoregulatory contractile aspect as seen and measured in the veins and arteries is missing or not observed. The microscopic size of capillaries (approximately 8μm) makes them less amenable to in-vivo observation, measurement or modeling of their structure. Their distribution en-masse, as vascular beds, has been studied, so too have many specimens from enucleated or cadaver eyes. These in vitro studies provide many details for the mechanisms by which diabetic retinopathy develops.

The retinal capillaries are composed from the interior to the exterior of the following three layers; the endothelial cell layer, a basement membrane and the exterior pericytes (or mural cells). Like the vasculature of the brain, the retinal vessels specialise in maintaining (the seemingly-contradictory requirements) of a high oxygen availability and the blood-retina barrier. This specialisation is marked by the endothelial cells forming the inner vessel wall.

The endocelial cells are organised in a layer of unfenestrated tight-junctioned cells that form an epithelium with no gaps through which red blood cells and other blood proteins, such as albumin, can pass. These cells rest on a basement membrane which contains collagen, laminin and proteoglycan which further prevents gross diffusional transport of blood solutes and particles. The opaque nuclei distributed approximately once every

other nulciec diameter visually characterise the pericytes which surround
the basement membrane. These details are only resolved using electron and
or light microscopy, and were gained from descriptions from the following
texts Ishibashi and Inomata (1993), Forrester and Knott (1997), Cunha-
Vaz (1997), Provis et al (1997) and Patel (1995). No anatomical canon for
the retinal circulatory system was found.

1.4.2.4   Distribution retinal capillaries

There are three distinct volumes bereft of capillary networks: the foveal
avascular zone, the more external retinal layers and the volume externally
concentric to the arterial vessels (Michaelson, 1948). Otherwise the retinal
capillaries are distributed over the entire retina. Their density reaching a
peak in the peri-foveal, macular and optic disk areas. Toussaint and Danis
(1970) demonstrated this by measuring the cumulative length of capillaries
for areas at various distances from the central retina. The capillary den-
sity distribution over the posterior hemispheric retina has a symentry line
running horizontally through the optic disk. Vertical symmetry was not
present; the temporal aspect, containing the macular area, having a denser
capillary network than its nasal counter part.

1.4.3   Pathology and aetiology of diabetic retinopathy

Malfunction of the blood retinal barrier causes early DR. These causes
have yet to be elucidated from the molecular biology of the retinal phys-
iology. These chemical changes eventually induce permeability, surface-
adhesive, structural and metabolic changes to cells of the retinal capillary
wall. This abnormal function is brought on by the toxicity of high glucose
concentrations (hyperglycaemia) as well as the toxic effects of by-products
from stressed metabolic pathways. The affected proteins have structural,
catalytic or mediator functions; thus there are diverse physical manifesta-

tions of early diabetic eye disease. These effects are believed to be compounded by the lack of a lymphatic or extra-cellular fluid drainage system. These small symptoms in turn give rise to more serious manifestations of DR as the retina attempts to compensate.

Auto-immune or viral destruction of the pancreatic B cells, leading to insulin deficiency, causes Type I diabetes (Forrester and Knott, 1997). The causes of Type II diabetes are not as clear. Poor action and production of insulin are suspected. The retinal consequences of both these diabetes types are similar. Any variability between the consequences arises out of the different time of diabetes diagnosis. Type II diabetics often have subclinical disease for many years before receiving health care, whereas the symptoms of type I diabetics are acutely manifested in other organs.

Among the myriad of consequences of hyperglycemia is the increased viscosity of diabetic blood, which is thought to exacerbate the hyperglycaemic damage to the endothelial cells and basement membrane. The appearance of fenestrae and the lack of tight junctions influence the permeability of the vessel walls (Ishibashi and Inomata, 1993). This toxic or biochemical insult can lead to complete endothelial and pericyte cell loss, so that the capilaries are acellular or composed of the basement membrane alone (Stitt et al., 1995).

The pericyte contractile cells, being only a partial structural component of the capillaries, are believed be regulatory in nature. Their loss combined with the loss of smooth muscle mass in the larger retinal vessels may increase the haemodynamic pressure leading to the distension and ballooning of capillaries also known as micro-aneurysms or MA. The damaged endothelium attracts and adheres the mobile and inflamatory cells of the blood. This leads to capillary blockage, MAs may represent abortive new vessel growth attempting to re-canalize or re-perfuse the blocked vessel (Forrester and Knott, 1997). MAs represent one of the first clear and easily perceptible changes in micro-vascular morphology. The consequences

of pericyte loss, endothelial cell damage and MA formation (potentially coupled with impairment of RPE function) are the starting point for the development of DR.

1.5 Progression of diabetic retinal pathology

The changes in micro and macro-vascular function and structure can lead to reduction or total loss of visual acuity. The progression, consequences and manifestations of the break down of the blood-retina barrier are different in each individual. A healthy diabetic retina may develop none, some or all of the lesions typical of early retinopathy. Lesions such as MA, seemingly so integral to the development of DR may be present, absent and or undetectable in early DR. Yet despite these irregularities the terminology used to describe the status and stages of DR is linear; starting with a normal or healthy retina, and ending up with a phase of significant scaring and retinal tissue loss. Examples of such scales are published by Klein et al. (1989a), Feman et al. (1995), ETDRS Research Group (1991b) and Davis et al. (1998). In between these two extremes there are usually two general categories describing the progression of DR. These two categories are Non-Proliferative and Proliferative DR, or NPDR and PDR respectively.

The progression of diabetic retinal pathology starts with the small and local lesions described in the aetiology of DR. Lesions such as Microanuerysms, small intra-retinal haemorrhages and calibre distortions called Intra-Retinal Microvascular Abnormalities or IRMA are typical of mild NPDR. These lesions can be seen in Figures 1.7, 1.12 and 1.8 respectively. Capillary blockage and capillary dropout may lead to larger areas of ischeamic retina. Focal areas of non-perfusion or infarcts lead to accumulation of axoplasmic

fluid, leading to soft exudation. These are diffuse white or greyish smears, also called cotton wool spots. A single soft-exudatative lesion is presented in Figure 1.10. Other lesions typical of mild NPDR are hard exudates; where failure to reabsorb, or the accumulation of, the leaked lipo-proteins present small, well defined and bright patches (See Figure 1.9).

At this point, further deterioration of the blood retina-barrier would take DR from a mild non-proliferative state to moderate or severe state. Moderate NPD is characterised by extensive intra-retinal haemorrhaging and a poorly perfused retina within which the venous vascular walls may distend. The distension alters the diameter and occasionally the path of the vessel. In other words, the vessels become tourtous-of-path or beaded-in-appearance as shown in Figure 1.14. Finally vasculogenesis or neovascularisation attempts to reperfuse the retina, a hall mark of PDR. These tufts of microvessels are fragile and grow into the interface zone between the vitreous and the retina (Figures 1.15 and 1.16). If this zone expands, as it does with age, the new vessels that straddle it from retina to vitreous, may rupture and haemorrhage. This type of haemorrhge is pre-retinal, its non-retinal nature can be appreciated in Figure 1.13. After Neo-Vascularisation Disk (NVD) or Neo-Vascularisation Elsewhere (NVE) scar tissues arise in an attempt to heal and clear the haemorrahge. These scars can contract and detach the retina with consequent damage to the retina. In total there are twenty-two retinal and preretinal lesions which have been monitored in studies assessing the progression and treatment of DR (ET-DRS Research Group, 1991c).

Among these lesions, macular oedema (MO) is most notable for its early and asymptomatic development (Klein et al., 1989b; Klein et al., 1995) and its successful treatment (ETDRS Research Group, 1985; ET-DRS Research Group, 1991c). Increased vascular permeability results in the accumulation of plasma-like fluid. Retinal swelling may also occur in the absence of significant extravasated fluid, in this case the cells of the

matrix of structural cells engorge or swell. Unlike the other lesions there is less of a progression or development of MO; the severity being defined primarily by location, and only to some extent by the volume of the swelling or fluid. It is the involvement of the macula in oedema which is the critical factor. Macular oedema is hardly perceptible via non-stereoscopic images yet, hard exudates may be an indicator of MO, specially if circularly distributed. Unfortunately the strength of the relationship between the presence of MO and hard exudation is unknown.

1.5.1 Definition and grading lesions and DR

With the exception of MO, the pathological development of the injuries is mirrored by the grading schemes used to assess the stage of DR. The most repeatable of the grading schemes, and one which best approximates a measurement or metric of the disease, is based upon the Diabetic Retinopathy Study grading scheme (DRS Research Group, 1981). This has subsequently been refined by the Early Treatment Diabetic Retinopathy Study (ETDRS Research Group, 1991c; ETDRS Research Group, 1991b; Davis et al., 1998). The photographic protocol, uses at least seven image pairs of each eye. Pairs of images are taken to provide stereopsis so that lesions which are outside of the retinal plane can be evaluated with depth perception, e.g. MO, drusen, NVD and NVE. It has been demonstrated in other human medical-image interpretation modalities that inter and intra operator variability is minimised when an explicit protocol is used (Peli, 1993; Scott and Palmer, 1993; Scott et al., 1996).

In DR classification these explicit protocols are provided by the ETDRS, and represent the best quantification of diabetic retinal pathology available. However the good detail and magnification associated with seven small (30°) fields of view datasets means that the 'gold standard' classification of DR is resource intensive. Alternative grading schemes, aimed

at screening or taking advantage of new imaging technology exist. The effectiveness of these classification protocols is tested against the seven-field colour stereoscopic fundus photography of the ETDRS. For example protocols aiming to: conserve resources (Aldington et al., 1995; Klein et al., 1986), remove the need for reference images (Feman et al., 1995), or take advantage of digital imaging (Aiello et al., 1998a) are all evaluated against the seven field stereoscopic standard.

In the latest revision of the ETDRS photographic grading scheme (Davis et al., 1998) eyes from the non-treatment arm of the clinical study were monitored for progression to PDR. Subsequently these eyes were divided into six categories of retinopathy-severity levels. Each severity level representing of the odds of developing sight threatening PDR. With the exception of haemorrhages, the lesions graded in the lower three levels (no DR up to mild NPDR) are not used in the last three. The lesions graded that contribute to these three lower levels are MA, hard and soft exudates.

The three most servere DR levels (4–6) are determined by the extent or area of lessions represented in Figures 1.12 through to 1.16. Individual levels are defined in combinations of lower levels. For example, there are two categories of moderate NPDR, between mild and severe NPDR. The more advanced of the two moderate NPDR, is combination of two lesser presentations of moderate NPDR.

The development of new vessels heralds the severe NPDR and PDR. With PDR being the first or original level at which photocoagulation therapy was demonstrated to prevent further vision loss (DRS Research Group, 1987). The point at which screening aims to intercept this harmful progression lies well in advance of this PDR stage. Ultimately the severity and extent of the lesions rests upon the comparison of the seven standard fields to reference photographs or against verbal descriptions, but note that DR is graded in terms of risk of developing PDR and not progression to the next level, de-emphasising the sequential nature, or progression, of DR.

Macular oedema is graded separately even though it is inherent in the development of DR. It is generally graded into two categories: 'clinically significant' MO or not clinically significant MO. If the macular oedema or (if present) its associated hard exudates are estimated to be within 500μm of the fovea, or if an area the size of an optic disk (OD) or larger of oedematous retina falls within one OD diameter the MO is deemed clinically significant. Otherwise any swelling is not clinically significant and below the threshold for which treatment has not been demonstrated to be effective.

1.6   Summary of Introduction

Thus far in this thesis, the medical, biological and some social contexts of diabetic retinopathy have been presented. In summary; DR is the largest remaining single cause of blindness in the working populations of the western world. The prevalence of persons at risk is high, with up to six percent of the population being at risk and in need of an annual fundus examination. Not only does screening provide a well established prophylaxis for sight-loss but it is also 'self financing' from a governamental perspective.

The implications of medical, clinical and epidemiological features of diabetic retinopathy on image processing of diabetic fundus images are important. For instance, the instruments used in screening determine the image characteristics; in terms of size, placement and quality. The screening protocols, clinic type and population determine the case mix and influence the prevalence of ocular co-morbitity. Together these factors characterise the data on which a potential system is developed and more importantly on which it is evaluated.

Unfortunately the vascular and oedematous lessions have varied prognostic value, both in isolation and combination. Their definition, detec-

tion, assessment and ultimately prognostic value is still unclear with respect to many of screening-imaging modalities considered. The screening recommendations for the number of fields, the field of view, need of stereopsis and even imaging modality conflict with the assumptions made in the financial models (Dasbach et al., 1991; Javitt et al., 1994). Digital image acquisition and the associated grading methods has not been demonstrated, despite its subjective convenience, to be a suitable imaging modality.

In the absence of clinical trials investigating the effectiveness of digital screening there is contradictory advice regarding screening methods (American Academy Ophthalmology Prefered Practice Patterns Committee Retina Panel, 1998; Aldington et al., 1995; National Screening Committee, 2000).

With doubt surrounding the basic assumptions of digital screening; i.e. that the lesions detected will translate into (cost) effective diabetic ocular-health-care, the author decided to concentrate on the segmentation of the retinal vessels. An issue of diminished contention within DR image analysis because even though the vessels are inherent to DR grading they can be argued to be less sensitive (but not immune) to, or independent of: choice of grader/expert, resolution, field of view, grading protocol, transducer types, display technology, stereopsis, underlying population demographics and financial implications.

Subsequent chapters will present a review on, and two methods for the segmentation of the retinal vascular arbour in reflectance images of the retina (retinographs).
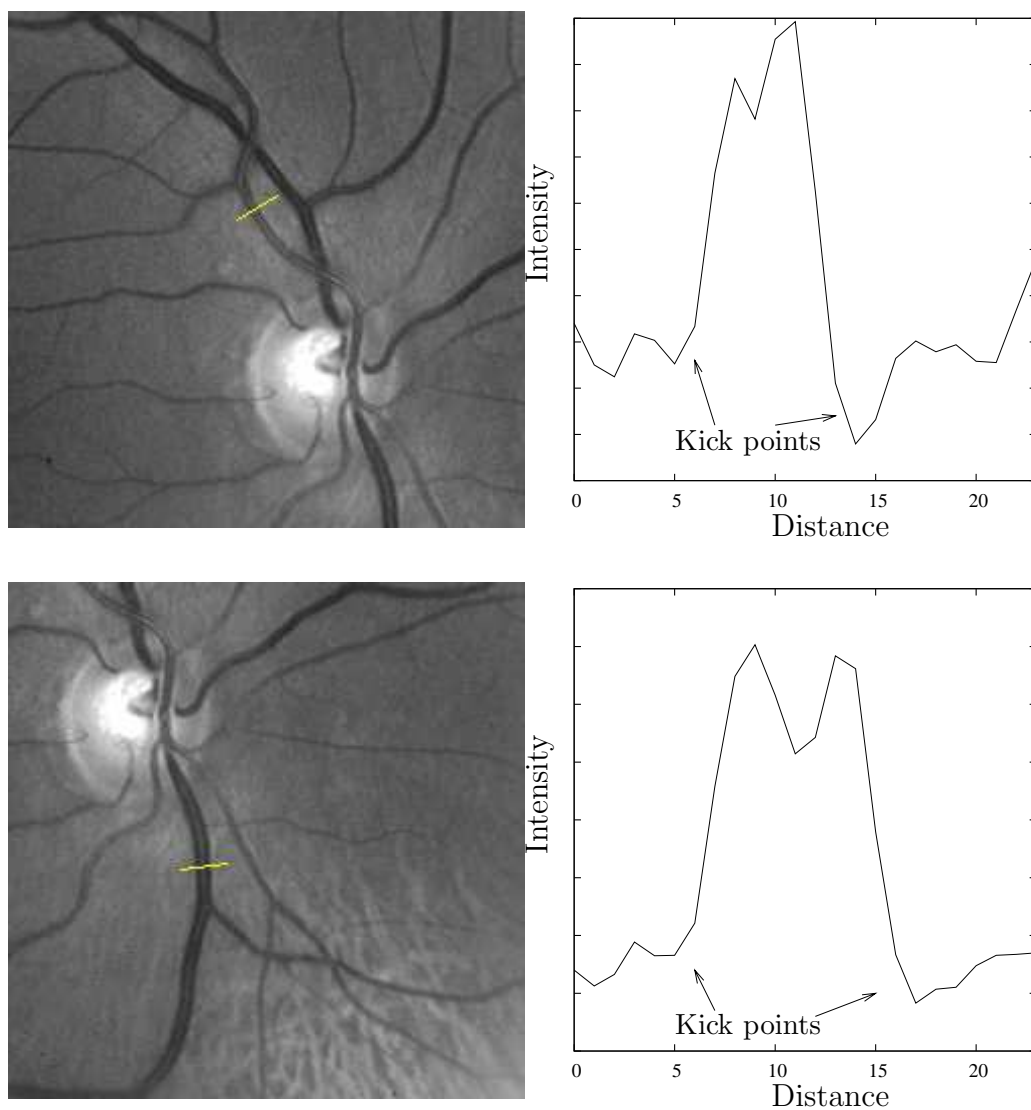
Figure 1.5: Inverted intensity profiles illustrating kickpoints at the suspected vascular lumen-to-wall interface
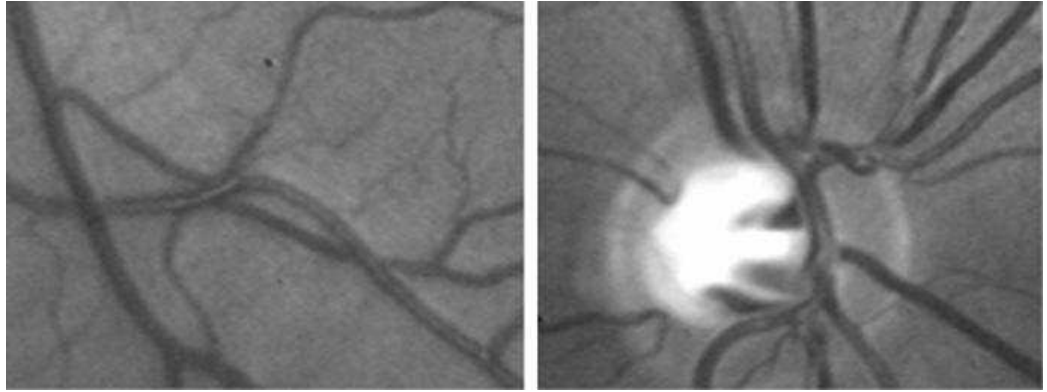
Figure 1.6: Monochrome fundus images showing twisted and overlapping vessels
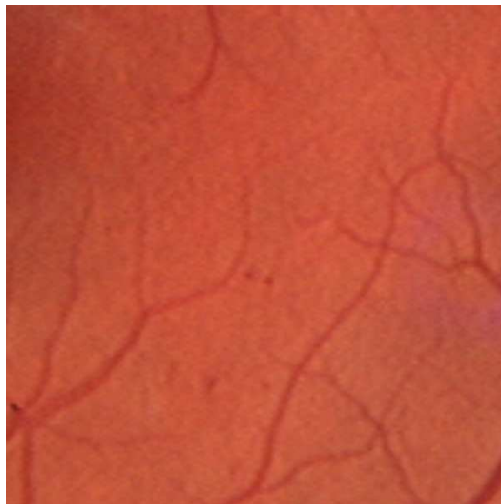
Figure 1.7: Microaneurysms, Small and punctuate red dots
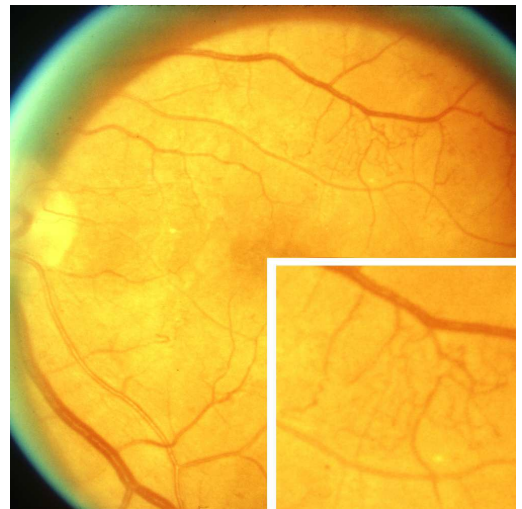
Figure 1.8: IRMA, Irregular calibres and tortousity of the smaller retinal vessels
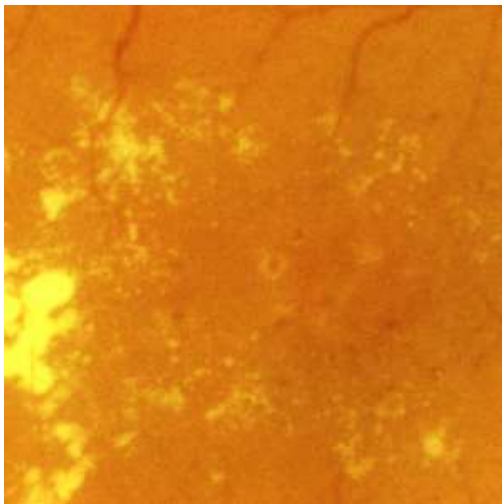
Figure 1.9: Hard exudates, areas of lipo protein mal-absorption, yellowish well defined contours.
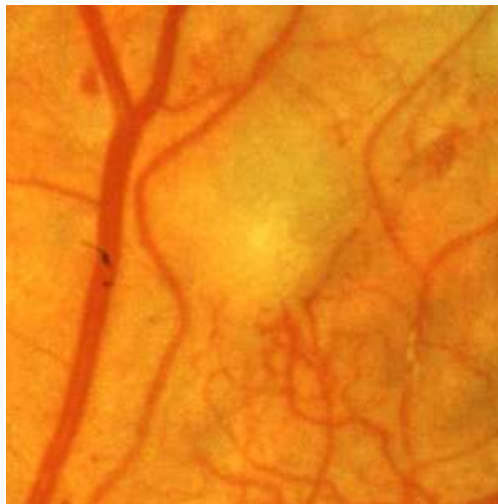
Figure 1.10: Soft exudates (SE), areas of axoplasmic accumulation
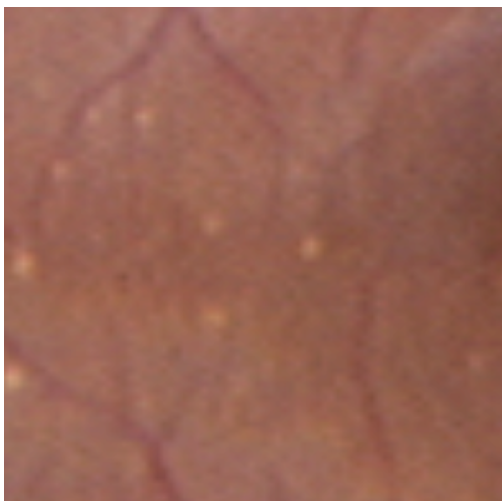
Figure 1.11: Drusen, Deep, whitish lessions more regular in outline than exudates
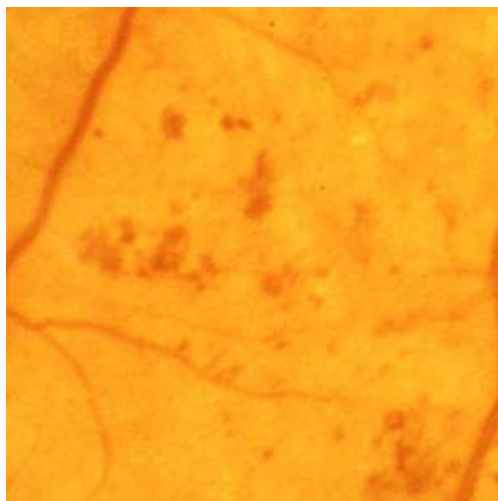
Figure 1.12: HAemorrhages (HA), dispersed red blotches with irregular perimiters.
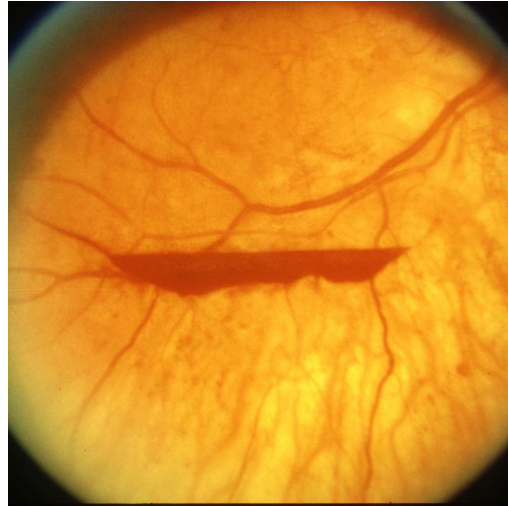
Figure 1.13: Pre-retinal Haemorrhage, extravasated blood between retina and vitreous ETDRS Std. Photograph 13
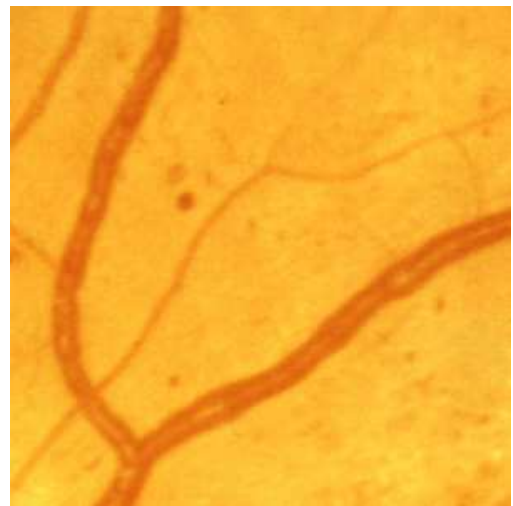
Figure 1.14: Venous beading, irregular distension of large and or smaller retinal vein calibres, excerpt of ETDRS Std. Photograph 6A

Figure 1.15: Neo Vascularisation Elsewhere, (NVE) Growth of capillary like new vessels on fundus but not OD

Figure 1.16: NeoVascularisationDisk, (NVD), Development of microvessels into the vitreo-retinal interface at OD

This page intentionally contains only this sentence.

Chapter 2

Retinal vessel image-processing review

The advent of digital images has driven the automatic identification and segmentation of the anatomical landmarks and pathology of the human retina. This chapter will provide an overview of the majority of peer reviewed work based on the analysis of reflectance fundus-images for the segmentation of retinal vessels. The review will encompass and update sections from previous work (Appendix A) undertaken in preparation for the research of subsequent chapters of this thesis.

The following sections start with a short look at the digital images themselves. This is followed by a summary of published research for the identification of the optic disk. Prior to the review of vascular segmentation methods, a section is dedicated to vascular profiles and the distribution of the vessels over the retina. Peer reviewed work on vascular segmentation is then split into four sections; local methods, matched filters, profile tracking and finaly a section neural network based methods..

2.1   Images: acquisition and resolution

The healthy tissues and pathological manifestations of the retina vary in size. The smallest vessels and their pathology (IRMA) barely larger than a capillary ($\approx 8\mu$m). The largest features such as pre-retinal haemorrhage and the optic disk are 1600μm or larger. The details of the anatomy and lesions that are recorded by an image are largely defined by the resolution. The capacity to resolve, or to spatially distinguish two objects, is a critical measure which is equally applicable to images recorded on analog media, such as photographic film, or those directly digitized from transducer signals.

A monochrome digital image can be represented as a two dimensional array with each cell or picture element (pixel) holding a value representing the intensity at that point in the image. This intensity is also called the gray level and is usually limited in resolution by the number of bits (or bytes) that represent it. The spatial resolution of a digital image is given by: its area in pixels, their size and the field of view. For a fundus image recorded on photographic film and through clear ocular media, the resolving power of the whole system is limited by the film and its development. A substantial ammount of research has validated the 'measurement' of DR using film. Grading using film based photography still forms the technical (and clinical) reference by to which other imaging systems are compared. It is now briefly described.

The spatial resolution of typical transparency film used for imaging DR, ranges between 1 to 100 cycles per mm for a response above 10 percent[1] Film resolutions of approximately 4000 pixels squared have been quoted by (Schalkoff, 1989). This figure is of the same order of magnitude as the 50 to 130 lines per mm, or 4600 lines per 35 mm quoted by the

---

[1]Data obtained from Kodak, Agfa and Fujifilm websites for Ektachrome, Agfachrome and Fujichrome transparency films.

Agfa data sheets. Film based evaluation of DR may use a relatively small
or narrow field of view, for example; during 35mm film-based stereoscopic
evaluation two 30° field of view images are usually used. Lesions of the
retinal capillaries such as microaneurysms and Intra Retinal Microvascular
Abnormalities (IRMA) require this resolution for unequivocal assessment.

Digital-image acquisition in contrast is usually over a larger field of
view; with 45, 50 or 60 degrees being typical. Images are usually rectan-
gular and have the long dimension ranging from 600 to 1024 pixels. Dig-
ital images thus represent larger fundus areas at lower resolutions. The
film-based reference imaging protocol also uses stereoscopic pairs which
the interpreter could integrate to their advantage. Digital transducers have
linearity and sensitivity advantages over photographic film. The fidelity of
reproduction and ease of transmission means digital imaging will become
an attractive alternative to 35mm film based methods once the resolution
approaches that of traditional film.

The National Screening Committee recommends that new digital sys-
tems should have a resolution of (1365 x 1000) capable of discerning cap-
illary sized pathologies such as NVD, MA and IRMA (National Screen-
ing Committee, 2000). Favorable evaluations of grading methodologies
based on digital images have been published: from a screening perspective
(Young et al., 1997; Kerr et al., 1997; Henricsson et al., 2000; Robbins et al.,
2001; Shiba et al., 2002) and from a grading perspective (George et al.,
1997; Aiello et al., 1998a; Fransen et al., 2002). All of these studies used
visualization resolutions below the acquisition resolutions recommended
by the National Screening Committee, and only one (Aiello et al., 1998a)
considered stereopsis. None unambiguously stated the spatial resolutions
represented by the sensors employed.

In summary, the spatial resolutions that are used can be roughly esti-
mated to from the width of the vessels or the average size of the optic disk
to give an resolution of the order of 10μm per pixel, see image 1.5.

2.2 Image pre-processing

After acquisition, corrections for uneven illumination or image quality may be made. Also adjustments to regions of interest or the need for re-sampling may be evaluated. Collectively these operations are termed image pre-processing. Other factors and artefacts include patient movement, poor focus, bad positioning, reflections, disease opacity or inadequate illumination. These factors and artefacts will cause a significant proportion of images to be of such poor quality so as to interfere with analysis.

The sphericity of the eye is a significant determinant in the intensity of reflections from the retinal tissues. It will compound artifacts such as the circular and crescent shaped low-frequency contrast and intensity changes arising from the interface of the anterior and posterior ocular chambers. In approximately 10% of retinal-images artifacts are significant enough to impede human grading (Kristinsson et al., 1997; Klein et al., 1989b; Liesenfeld et al., 2000). A similar proportion was found to be of inadequate quality for automated analysis (Chaudhuri et al., 1989a; Walter and Kllein, 2001).

Preprocessing of the fundus data either removes or flags the aforementioned interferences. Images with quality issues are either, purposely included, excluded after some tests or most frequently, issues concerning quality are silently ignored. Two studies assess the quality of images prior to further analysis. Both Lee and Wang (1999) and Gagnon et al. (2001) use intensity histograms, or histograms of feature metrics in the latter case, derived from ideal images. Deviance from these reference histograms is then assessed via correlation or chi-squared respectively. To demonstrate robust and flexible operation poor quality images are included in the test datasets (Hoover et al., 2000; Jiang and Mojon, 2003).

Pre-processing steps, in a similar fashion to quality checks, are: purposely excluded, purposely included, and in cases where it may be a con-

tributing factor and arguably not integrated into the main algorithm, often
silently ignored. Typical pre-processing operations include masking ROIs
(Hoover et al., 2000; Gagnon et al., 2001; Englmeier et al., 2002) and
noise suppression (Leandro et al., 2001; Ye and Zheng, 1995). However
the most common preprocessing step is performed to counter the irregular
illumination of the fundus.

Okazaki and Tamura (1983) approximated unequal retinal illumination
as spherical. Their surface fitting algorithms was successfully applied to
the image after setting a range of thresholds. The average centre of these
binary images provided enough terms to apply the least squares method to
find the radius and vertical offset.

A simpler approach was used by Lee (1992) and others (Shin et al.,
1999; Ege et al., 2000), who all used a 31 pixels squared median-filter as
a smoothing operator to obtain a representation of the image detailing
only the low frequency image content. This image was then subtracted
from the original to remove the effect of unequal or irregular illumina-
tion. Smoothing with Gaussian kernels is also employed in this manner
(Walter and Kllein, 2001). Other background changes introduced during
film-digitisation have been removed by subtracting (Frame et al., 1997a;
Frame et al., 1997b) or dividing (Cree et al., 1997) the uneven illumi-
nation known as the flood image. Adaptive local contrast enhancement
was applied by Sinthanayothin et al. (1999) and Osareh et al. (2002) who
rescaled pixels to the full intensity range, depending on the local mean and
variance. Most studies ignore the blue and red information of the images
and rely on the green plane to obtain the best contrast (Chaudhuri et al.,
1989b; Goldbaum et al., 1989; Hart et al., 1997; Kochner et al., 1998;
Lee et al., 1999; Martínez-Pérez et al., 1999; Zahlmann et al., 2000; Gagnon et al.,
2001; Walter and Kllein, 2001; Zana and Klien, 2001; Englmeier et al.,
2002; Jiang and Mojon, 2003).

2.3   Locating and segmenting the optic disk

The Optic Disk (OD) is an important structure to identify and locate as
the retinal vascularture originates from, or converges towards it. The OD
also acts as calibrating structure, giving a frame of reference by which to
size other anatomical structures (Wilson et al., 1988; Kristinsson et al.,
1997). It is the brightest feature of the healthy fundus, approximately cir-
cular, and is easily observable in many of the figures. The OD lacks a vis-
ible capillary supply and appears paler due to the lack of an underlying
pigmented layer.

The contrast between the emerging translucent blood-filled vessels and
the white background has been successfully exploited (Sinthanayothin et al.,
1999). A local measure of pixel intensity variance, of window size equal to
that of the OD (80 pixels) was swept over the image. The peak variance
measured was successful in locating the OD in 111 out of 112 test images.

Instead of using the variance, the average brightness value alone, has
also been used to locate the OD (Lee et al., 1999; Tamura et al., 1988).
A combination of many features, such as edge concentration, colour and
brightness were weighted by Katz et al. (1988) to highlight regions with
OD like properties. The brightness of the OD region was also used by Wal-
ter and Kllein (2001) who located OD centres via the mean location of an
area-constrained thresholding, and the OD contour by using the watershed
algorithm on a morphologically opened (see Section 3.1.3.1) and recon-
structed image. This approach was successful in 27 out of 30 images. It
failed on 3 images because of low contrast, saturated red plane, images.

Chaudhuri et al. (1989a) used local measures of vessel convergence and
orientation combined with a matched filter for bright circular objects to
locate the optic disk. This resulted in OD localisation of 31 optic disks in
35 predominantly pathological images.

The convergence of the retinal vessels upon the OD was more exhaus-

tively exploited by Hoover and Goldbaum (1998) and Akita and Kuga (1982). The later study identifies the order by which the vessels branch from the OD origin and measure the intensity of the fundus around this ranked list. The brightest pixel in the neighbourhood of the highest ranked, or root vessel was then identified as the OD centre. Hoover and Goldbaum (1998) used elongated and smeared versions of vessel segments. The diffused vessels accumulated and overlapped in areas of convergence such as the optic disk.

The Hough transform has also been used to locate the optic disk (Kochner et al., 1998). The two main vascular arcades were segmented to obtain points to which an ellipse was fitted. At the apices of this ellipse the OD was prospectively located using matched filters (see section 2.6.3) to find the major vessels leaving the OD. The response of the matched filters will be maximal if radially distributed vessels are present within a bow-tie shaped region of interest (ROI) centered on the presumed location of the optic disk. A Hough transform for circular feature detection has also been employed (Yulong and Dingru, 1990). The edge points were derived from the whole image using a morphological background subtraction on a gaussian blurred image.

Wang et al. (1990) fitted circles to the OD by successive approximation. Their approach started with thresholding so that pixels were iteratively rejected from the resulting binary image by discarding 15% of pixels not within a shrinking circle centered on the mean location. If after shrinking the origin of the circle could not be moved so as to reject a further 15% of pixels not with in the circle, the OD was taken to be identified. A similar fitting approach using contour snakes was employed by Mendels et al. (1999), the authors report qualitatively accurate results.

2.4   Vascular profiles and distribution

A few details on vascular crossections and their distribution on the retina
are now given. The macroscopic structure of the larger vessels has lent it-
self to modeling. Many representations of vessel cross-sections have been
suggested for the automatic identification of the vessels during tracking or
matched filtering. However the earliest use of vessel profiles is in vessel cal-
ibre measurements to link objective anatomical features to retinal disease
or physiology. The preferred method for this width measurement is to use
the Full Width at Half Maximum (Brinchmann-Hansen and Engvold, 1986;
Newsom et al., 1992; Chen et al., 1994; Suzuki, 1995; Dumskyj et al.,
1996; Lee et al., 1997). The FWHM method is resilient to poor focusing
but not as accurate as identification of blood-column and vessel-wall inter-
face (Rassam et al., 1994; Patel, 1995).

In the concentric ring model (used to accurately measure vessel di-
ameters) the light absorbance of a retinal vessel composed out of a ves-
sel wall and lumen is modeled using Lambert's law. The output intensity
I for an input intensity I_0 after traversing a length of blood filled lumen
l_l, with absorption coefficient α_l was given by $I = I_0 e^{-\alpha_l l_l}$. The frac-
tion of light I/I_0 passing through, or transmittance T, is then given by
$T = I/I_0 = e^{-\alpha l}$. The same transmittance calculation is performed for
vessel wall using α_w and l_w so that the total transmitance for the wall and
the lumen is $T = e^{-\alpha_w l_w + \alpha_l l_l}$. The leftmost graph of Figure 2.1 illustrates
the expected transmitance. The half lengths of the lumen and annular wall
structure are given by

$$
\begin{aligned}
l_l &= \sqrt{r_l^2 - x^2}, r_l > |x| \\
l_w &= \sqrt{r_w^2 - x^2} - l_l, r_w > |x|
\end{aligned}
\tag{2.1}
$$

where $\alpha_w = 0.65\alpha_l$, and $r_{\{l,w\}}$ gives the radius of the wall or lumen.
Patel (1995) assumed widths of 100μm for retinal veins and arteries and
wall widths to be 15μm and 10μm respectively.

Hammer et al. (1997) presents an elaborated version of the concentric ring model by simulating the many possible pathways of the light to contribute to the vessel profile. Including those backscattered by the many tissues behind the vessels such as the RPE, choroid and sclera. They conclude by suggesting that a 6 degree polynomial is sufficient to fit their simulated profile, alas without providing the w_i terms for the polynomial. The simulated profile included a strong reflection of the illuminating light from the lumen-wall interface where this was near-normal to the sensor.

Annular rings models for the retinal vessels have the advantage of producing the kick points, or skirts which may be observed in high resolution profiles of (Brinchmann-Hansen and Engvold, 1986) and in the less detailed images (Figure 1.5). The clear inflections modeled in the leftmost graph of Figure 2.1 are also seen in data from Zhou et al. (1994) (Figures 2 and 4), in Figure 2 of Tolias and Panas (1998) and Chen et al. (1994) and Figure 8 of Delori (1988). The vessel kick points are over represented in the schematic representations of Suzuki (1995). These kick-point features will become important because vascular distensions have been shown to precede the development of MO (Kristinsson et al., 1997); potentially providing a non-stereoscopic diagnosis or indicator of MO for which depth perception is usually necessary.

The vessels in Figure 1.5 also show how the larger arterial vessels may have a bright streak running along the apex (near normal presentation). The accompanying intensity profiles also demonstrate a clear inflexion. Gaussian, concentric ring, and elliptical (Jasiobedski et al., 1993) profiles lack a central light-reflex. As a result, Iliasova et al. (1998) and Gao et al. (1997) have proposed fractional-rational (Equation 2.2) and difference of Gaussians (Equation 2.3) to better approximate the relationships between the observed intensities and distance across the vessel. Their profiles are
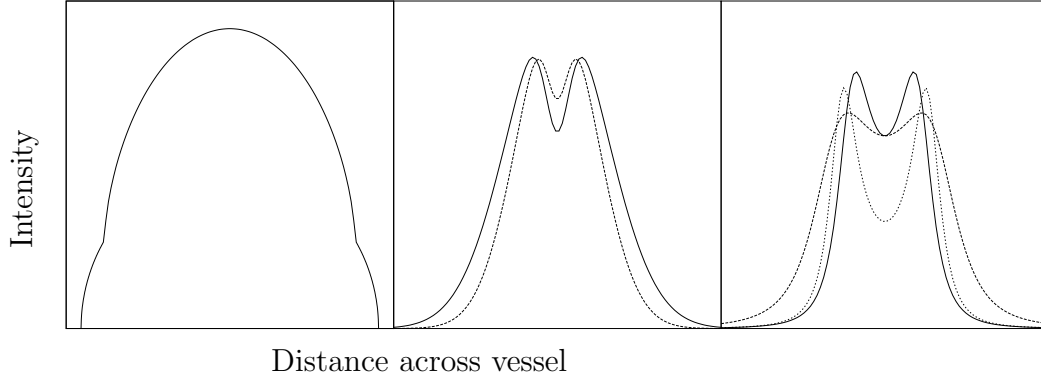
Figure 2.1: From left to right, concentric ring, fractional rational and difference of Gaussians. These profiles have been suggested to model the profiles of retinal veins and arteries. Compare the shape of these theoretical profiles with sample profiles shown in Figure 1.5

presented in Figure 2.1.

$$T = \frac{A}{(x - x_0)^4 + a(x - x_0)^2 + b^2} + C \qquad (2.2)$$

$$T = A + B_0 e^{-\left|\frac{x-x_0}{\delta_0}\right|^2} - B_1 e^{-\left|\frac{x-x_0}{\delta_1}\right|^2} \qquad (2.3)$$

where A, B, C and δ are the parameteres conrolling the shape of the profile, and T is the transmittance. These two later profiles are suggested, but not demonstrated, to have greater fidelity to observed profiles.

In contrast to all of these profiles the observed profiles across real vessels segments have irregular wall thicknesses and variable lumen shapes. Profile fidelity could be enhanced by incorporating the peri-arterial capillary free zone. Sometimes this is perceptible as a bright zone flanking the larger arteries. The vessel profile also changes shape as the vessel divides; at the bifurcation point itself, and beyond in subsequent branches where the kick points are too small to be reliably sampled. Parenchymal dips, twisting of vessels, and numerous Arterio-Venous crossing phenomena (Akita and Kuga, 1980) are some of the vessel states where the profile

will also be distorted. For segments free from these interferences a single profile has been found to be too variable. This has been compensated for by: using a length of segment (Dumskyj et al., 1996; Englmeier et al., 2002), averaging 12 (Newsom et al., 1992) or averaging 5 (Chen et al., 1994) profiles from the same segment.

2.5   Retinal vessel distribution

The distribution of the vessel segments, and their branching behaviour over the retina will now be described. This description is based on space-filling (fractal like) concepts or tree like grammars. It has long been observed that the vessels follow a self similar motif (Toussaint and Danis, 1970), and recently a model explaining the variation of calibre, segment lengths and the distribution of the vessels has been suggested (West et al., 1997).

The attraction of using fractal descriptions of the vessels is that a single number can provide a scale invariant description of how the vessels fill that region. This number is the fractal dimension. It has been used to investigate the patterns of neovascularisation. The fractal dimension, as measured in a region of interest (ROI) using the density-density correlation method, has been shown to change for a patient over the course of neovascularisation (Daxer, 1993a) and between normal and neovascularised retinae (Daxer, 1993b). Landini et al. (1993) measured the fractal dimension for whole retinal images of normal retinas, and for those with venous oclusion (Landini et al., 1995) and found no difference. However, the fractal dimension of connected components within subwindows did distinguish the pathological and normal states.

The retinal vascular arbour has been suggested not to represent a true fractal but better described as space filling lattice (Panico and Sterling, 1995; West et al., 1997). The scaling laws for this lattice have been de-

duced by West et al. (1997) in biological systems. They make three assumptions: a bounded capillary size, that tissues are perfused, and that 'the energy to distribute resources is minimized'. These assumptions are used to derive scaling laws describing the ratios in vessel calibre and segment lengths across branchings of the vascular system.

In pulsatile systems such as the human one they predict the lumen diameter (l) at branching level k varies so that:

$$l_k/l_{k-1} \approx \begin{cases} \gamma_{prevenule} & , k < 7 \\ \gamma_{postvenule} & , k \geq 7 \end{cases} \tag{2.4}$$

where γ is the segment length ratio.

A similar relationship is hypothesized to hold for the segment length. Interestingly this discontinuity in self similarity was independently observed in the fractal analysis of the normal human retina, where two gradients are clearly present in the self-similarity plot of Figure 2 in Landini et al. (1993), but the author is not aware of any further epirical tests.

Reconstructions or descriptions of the retina using tree-like grammars have focused on the branching or tree morphology of the vessels, and have not integrated the aforementioned scaling observations (Hesse et al., 1993; Kókai et al., 1999) which apply to segment lengths and calibres. Whilst ignoring length and width, the tree like grammars highlight the importance of branching angle and vessel-segment curvatures in recreating the vascular arbours. Kristinsson et al. (1997) have linked vessel segment lengths and calibre distensions to MO. Martínez-Pérez et al. (2002) have linked segment lengths and vascular calibre changes to hypertension.

2.6 Introduction to vascular segmentation

Automated analysis of vessel calibres and lengths requires the retinal vasculature to be identified and segmented. The measurement of the vascu-

lar morphology is not the only reason to identify and delineate the vessels. Planing and delivery of laser photocoagulation therapy, constructing image montages and assessment of the retina between vessels also requires the segmentation of the vessels.

The author has categorised the segmentation approaches into methods based on: local information, profile tracking, matched templates and neural networks. Local information is exploited by thresholding, curvature analysis or edge strengths. Profile tracking iteratively matches cross sections of the previous section to follow, or trace the path of the vessels over the fundus. Matched filtering can be considered to be brute force variation of profile tracking; where profiles are extruded into an extra dimension and prospectively positioned. Neural networks are also used for vascular segmentation, details on their architecture will be presented.

2.6.1 Local and region growing vascular segmentation

The retinal vessels are darker than their surroundings enabling segmentation by threshold. Pixel intensities above this threshold can be accepted as belonging to a vessel and those below discarded as background. Careful selection of regions enables local thresholding to work. For example, studies assessing venous beading (Kozousek et al., 1992; Gregson et al., 1995; Shen et al., 1991) only measure the variation in vascular calibre along the larger vessels, where the pixel intensity distributions are well separated from those of the retinal parenchyma. Gregson et al. (1995) chose the regional threshold as equidistant from the means of background and vessel pixels, which in turn were defined as top and bottom cumulative percentiles, the percentage was not given.

Simple thresholding has limitations, as it is confounded by pathology and unequal illumination. The variability is countered by enhancing the

pixels on a regional basis. The local gradient, curvature, connectivity, shape
and area are all used, often with constraints set using a-priori knowledge.
Akita and Kuga (1979) and some of Jiang and Mojon (2003)'s method,
use multiple thresholds together with subsequent area or shape based dis-
crimination to discard components too small, large or non-curvilinear to be
vessels.

Jiang and Mojon (2003)'s method is more elaborate than threshold-
ing alone. Initially the threshold is performed over a range, at each value
candidate areas are identified these are tested for inclusion. Succesful can-
didate areas are then OR'ed to produce the final classification. Additional
criteria for inclusion are based on contrast measures and metrics derived
from the euclidian distance of candidate pixels to the centre line of the
'blob'. The results presented are categorised according to image type being
classified; normal images are classified so that 92% of all pixels classified as
vessel actually are vessel. The remaining 8 % is mis-classified background.
Similarly 92% of pixels classified as background is background, the remain-
der are actually vessel pixels. In images with pathological features these
accuracies drop to 87%.

The local gradient of the image, in the form of first and second di-
rectional derivatives, have also been exploited for vascular segmentation.
Yu et al. (1990) and Katz et al. (1988) use the first derivative or Sobel
operator to find the magnitude and direction of the parallel edges of the
vesels. They use the opposing and parallel gradients along the vessel to es-
timate the location of the vessel midline. The second derivative, smoothed
by a Gaussian of $\boldsymbol{\sigma = 4}$ was investigated by Martínez-Pérez et al. (1999).
They used the eigenvalues of the second derivatives of each pixel to obtain
the maximum and minimum gradient irrespective of their orientations.

Morphological based methods (See Chapter 3) have also been applied.
Five studies have exploited morphological openings using linear structuring
elements (Spencer, 1992; Rambhia, 1999; Leandro et al., 2001; Walter and

Kllein, 2001; Zana and Klien, 2001). The morphological approach exploits
the three dimensional geometry and shape of the vessels. A linear, rod like,
structure is swept raster style at various orientations over the landscape
of retinal image. The maximum depth (or height) of the various ways this
structure fits, without intersection, into the valleys or ridges of the vessels,
is taken as output.

Mathematical morphology is also used by Ye and Zheng (1995) who
used a spherical structuring element that is swept or 'rolled' over the re-
lief created by the intensity peaks and valleys of the image. The sphere
does not fit into the valleys of the vessels but rolls over them. In effect the
background image is created from the patches or three-dimensional-arcs
that remain. This morphological approach also presents its results via sam-
ple images and textual descriptions. The best of among these subjective
descriptions-of-performance is provided by Zana and Klien (2001) who in-
clude the results of testing on images with noise, different resolutions and
pathology.

The difference in colour between the vessels and the background has
been exploited too (Akita and Kuga, 1982). They used analogues of hue
and saturation from the Unified Color Schema of the CIE (Commission
Internationale d'Eclairage). The vessels were clearly linearly separable from
the retinal background, although the number of measurements was limited.
The aim of the research was to identify veins and arteries separately, the
colour segmentation was insufficient for this. Lalonde et al. (2000) found
colour information to be insufficient even to distinguish the background
from the retinal vessels.

2.6.2   Profile tracking for retinal vessel identification

From a start point on a vessel, typically found by locating the centre of
the OD first, profile tracking iteratively scans for, selects and follows pixels

belonging to the retinal vascular arbour. The selection of vascular points is
accomplished using two or three dimensional matched filters. Fuzzy vessel
profile recognition and intensity based algorithms provide alternatives to
matched filtering (Tolias and Panas, 1998; Wang et al., 1990)

The two-dimensional profile tracking algorithms reviewed used Gaus-
sian or derivative of Gaussian filters. Profiles are available that represent
the vascular crossection with less divergence in width (Rassam et al., 1994),
however these are not as analytically simple to allow least square fitting
(Frame et al., 1997a) or for deriving indices (Tamura et al., 1988; Zhou et al.,
1994) to define vessel edge and thus width.

The profiles used are only apt in describing single vessel segments, but
vasculature branches and arterio-venous crossings are common. This presents
a special case for tracking. Frame et al. (1997a) use a measure of pro-
file skewness to create new starting points for the algorithm. In a similar
fashion increased vessel-width triggers sampling of a relatively large and
circular crossectional profile to which the filter is applied (Tamura et al.,
1988). The number of responses indicating whether branching, crossing
or a vessel end point has occurred. Can et al. (1999) use an asymmetry
measure of the template response to trigger new tracking. Prospective
scans alongside, or parallel to, the vessels which were segmented during
the first pass are then used to detect branchings or crossings (Liu and Sun,
1990; Kochner et al., 1998; Chutatape et al., 1998)

A tracking approach using an adaptive Gaussian profile, capable of
accommodating different vessels widths, is claimed to increase sensitiv-
ity (Gang et al., 2002). Digitised and digital images were tested with the
adaptive profile. In 1873 vessels the adaptive width algorithm erred 2.5%
of the time, the fixed width version 7.4%. Vessels were defined as 'longer
than 1/2 the disk diameter'. Tolias and Panas (1998) used a similar mea-
sure to communicate the ability of fuzzy c-means in tracking 'major' and
'minor' vessel segments. No false positives were detected out of 68 vessels,

however only 4 out of 17 minor vessels were detected. Major vessels were found in 50 out of 51 cases. Can et al. (1999) use two separate ranges of a Gaussian curve to track the flanks or skirts of a vascular profile, reducing the number of calculations, and thus, computational time to processes an image.

Englmeier et al. (2002) use an axially smoothed profile of the vessel itself for tracking. The maxima and minima of the first derivative of this natural profile are then used to find the vessel edges. The authors test the algorithm on images from various cameras and sources, with and without interfering diabetic pathology. They describe their results as satisfactory with the following provisos: no excessive tortousity, sufficient resolution and acceptable contrast and quality. A similar approach using the sign of the derivative on a single-profile, smoothed by a Gaussian kernel, has been taken by Tascini et al. (1993).

2.6.3 Matched filters for retinal vessel segmentation

Matched filters are extruded versions of their two dimensional counterparts. Matched filters or templates in contrast to the selective application used in vessel tracking are are convolved with an entire image. Matched filters are based upon similarity measures between an ideal and measured signal. The sum of products of the translated matched filter over every point of the image is usually taken as the filter's output. To obtain a good fit, the profiles of Section 2.4 are defined using a number of parameters; such as length, width, and orientation. Many other parameters are ignored. Nevertheless numerous potential matches need to be evaluated making this approach computationally expensive in the absence of dedicated hardware.

As retinal imaging varies in field of view and number of samples, parameters for the filters are thus resolution dependent. For the extruded and rotated Gaussian filters, parameter selection has not differed from the orig-

inal study (Chaudhuri et al., 1989b), with $\boldsymbol{\sigma = 2}$ (Goldbaum et al., 1989; Goldbaum et al., 1993; Goh et al., 1997; Hoover et al., 1998; Yang et al., 2000).

Linear filter and derivative of Gaussians filter shapes have been investigated (Wang et al., 1990; Poli and Valli, 1997). These filters consider computation time as an important design factor, and rely upon binary or integer arithmetic taking less clock cycles than floating point calculations. Wang and S.C. (1997) apply filters 15 pixel-squared, containing a line a single pixel wide, over 12 orientations. Poli and Valli (1997) settled for a standard deviation of one, in their binomial approximation of a Gaussian. Quantitive measures to characterise the performance of these studies are desirable, yet lacking.

Matched filters do not work in isolation and form part of an algorithmic chain. They provide a non-binary measure of correlation that still needs classification into background or vessel. Thresholds divide the range of intensity of the filter response into two classes which are applied locally or globally. Hoover et al. (2000) iteratively decrement the initial threshold (which is set to the matched filter response) so that a vessel blob is grown. This blob may be kept or discarded depending on tests of size, relative position to already marked pixels and absolute threshold value.

In this category, Hoover et al. (2000) has provided the only published objective results. They compared classification accuracy against a human operator, and reported a classifying 84% of the vascular pixels as vessel and 84% of of the background as non-vessel. Vascular segmentation algorithms currently operate at much (5 - 6%) higher accuracies.

2.6.4 Artificial Neural Network segmentation

The operation of a neural network is analogous to that of a matched filter. Both take subwindows of the image as input and return a probability mea-

sure as output. Two studies, both using the back-propagation algorithm, have detected (Gardner et al., 1996) and segmented (Sinthanayothin et al., 1999) the retinal vasculature. Detection entailed classifying subwindows as containing vessels or not, segmentation classified individual picture elements.

The neural networks require training examples, these examples were created by a human operator. Sinthanayothin et al. (1999) used 25094 labeled (pixel by pixel) examples each of 200 pixels in area. However it was not clear if they labeled ≈ 2.5 million pixels to create independent and non–overlapping examples or if they labeled an area which can generate 25094 examples whose input (but not output) areas overlap. The neural networks researched by (Gardner et al., 1996) used 20 by 20 pixel subwindows. Nine thousand of these subwindows were marked for neural learning validation. Generalisation assessment over 1200 unseen subwindows resulted in a sensitivity and specificity of 91.7%. Sinthanayothin et al (1999) achieved a sensitivity of 83% and a specificity of 91% in classifying picture elements.

2.7 Chapter summary

Vascular segmentation of the retina has been attempted using a wide variety of methods. Many of these on angiographic images, in which the vessel signals have been enhanced using contrast agents. However as increased image resolution and better focusing technologies become available the methods developed for these contrast enhanced images will be applicable to reflectance images.

Some successful strategies have been developed and tested to located the OD (Sinthanayothin et al., 1999). This is an important aspect for the tracking of the vessels.

It is hard to suggest which of the approaches successfully segments the vessels. Three of the methods present objective results (Sinthanayothin et al., 1999; Hoover et al., 2000; Jiang and Mojon, 2003). For predictable operation these methods should be considered in preference to others. Some studies (Zana and Klien, 2001; Englmeier et al., 2002) extend their results beyond the presentation of a sample segmentation by testing on images of lowered quality or with pathology.

Chapter 3

Morphological retinal vascular segmentation

3.1 Introduction

From the review in Chapter 2 it may be appreciated that many image processing techniques have been applied to the segmentation of the retinal vessels. The main drawback for the majority of this research is the subjective presentation of results. In this chapter a vascular segmentation-technique, based on mathematical morphology methods is presented with full results. The morphological methods rely on the shape and connectivity of picture elements, i.e. how they are connected to each other, to process the images. The performance of mathematical morphology, which has not been quantified to date, will be evaluated against human labelled ground truth images. This quantification unambiguously and explicitly conveys how the classifier, which in this case is a sequence of morphological operations, identifies both vessel and non-vessel picture elements.

Knowledge of the classification accuracy and the expected behavior over a varied dataset is necessary to demonstrate robust operation. Both the

author's and the external datasets over-represented pathology or problematic image acquisition (Hoover et al., 2000). As a result in many cases the vessels are obscured, this makes their identification erratic. Difficult and atypical data is necessary to unambiguously demonstrate the operation of the approach on all types of images that are likely to be seen in clinical scenarios.

The author chose to measure the efficacy of classification relative to that of a human grader because ground-truth of greater fidelity was not available. The goal of the image processing detailed in this chapter was to label all pixels into vessel and non-vessel classes as a human would. Human graders, both inside (Thomas Teng) and outside (A. Hoover, henceforth AH) this study labelled a total of 35 images. Each pixel in each image was thus classified into the vessel and non-vessel classes.

Unfortunately it was not possible to standardize labeling methods. The Hoover et al. dataset being published during the labeling process of the local images. Hoover et al. (2000)'s publication in conjunction with the public availability of the actual retinal images, represents the first and only study thus far to have shared their data in this fashion. In doing so Hoover, Kouznetsova and Goldbaum have enabled performance-evaluation of algorithms for the segmentation of the retinal vascular arbour with a greater degree of objectivity than possible before. The availability of this reference data is leveraged to prospectively assess the behavior of the proposed segmentation methods. The merits of this study in the absence of the AH data, would have rested upon a prospective evaluation on data withheld during development.

The following sections will describe the data and its subsets. Background information for the basic morphological methods used will be given and is followed by a description of the actual algorithm used. The methods used to summarize the segmentation performace of the retinal images is then presented. Finally the actual results and conclusions are presented.

3.1.1   Description of Hoover et al. images

The AH images, and reference data are the product of a study that also
attempts to identify the vasculature of the retina from reflectance fundus
images. The study assumed human grading to be accurate and consistent.
An assumption based on five double gradings performed by Hoover et al.
(1998). A subsequent paper reinforced these assumptions (Hoover et al.,
2000). In this study twenty ground truth images independently created by
the two graders achieved a 90% correct non-vessel reclassification rate, and
an 94% correct vessel reclassification rate.

3.1.1.1   Retinal image quality and the presence, or absence, of pathology

Ten of the twenty 35° images were chosen to include significant confound-
ing pathology so as to increase the rigour of the assessment. The remaining
ten healthy fundus images were captured in-focus and free of obstructing
artifacts. Images such as 4, 8 and 9 in Table 3.1 either were not in focus,
had opacities of the ocular media or were affected by imaging artifacts
other than those caused by pathology. These images were termed 'poor
quality'. The AH images with corresponding labels were publicly available
from the STARE project web space.[2].

3.1.2   Description of Royal Bournemouth Hospital images

These images were obtained two to three years after the digital fundus
camera came into routine use at the hospital. Patients for the 'joint retinal
clinic' as the imaging session was named were forwarded for retinal imaging

---

[2]STARE Project images can be obtained from http://www.parl.clemson.edu/stare/probing/{stare-images.tar,labels-ah.tar} Accessed 26th June 2001.

Image label	Image Number	Pathology Present	Poor Quality
im0001	0	●	
im0002	1	●	
im0003	2	●	
im0004	3	●	
im0005	4	●	●
im0044	5	●	
im0139	6	●	
im0319	7	●	
im0240	8	●	●
im0324	9	●	●
im0291	10		
im0077	11		
im0081	12		
im0082	13		
im0162	14		
im0163	15		
im0235	16		
im0236	17		
im0239	18		
im0255	19		

Table 3.1: Composition of Hoover et al. (2000) image data base.

by the Diabetes and Endocrinology Department. The images were captured using pharmacologic dilation of the pupils. The equipment that was used to capture these images consisted of a Topcon FRC 50X ophthalmoscope with a Sony 3CCD image transducer attached to Personal Computer running the 'ImageNet 1.53' image management software. All images went into long term storage on a HP Surestore Optical 2600FX. It is from this database that Tagged Image File Format images were retrieved. Non image data stored in the image tags was removed.

Fifteen images, 640 pixels wide and 480 pixels high, were selected and labelled. These images were of a variable field of view (20° to 50°) because of the fundus camera's analogue, or continuous, field of view selection con-

Image Number	Pathology Present	Poor Quality
0		
1		●
2		
3		
4	●	●
5		●
6	●	
7	●	●
8		●
9		●
10		●
11	●	
12		●
13	●	●
14	●	

Table 3.2: Composition of Royal Bournemouth Hospital (RBH) image data base, that was labelled by the author

trol. All types of fundus area are represented in the data set. Images centered on the macula, OD and perifery were included.

Ideally images would have been selected at random, however it is likely that results would be skewed by the preponderance of images without pathology or imaging issues. In the author's opinion, selection was necessary to ensure enough images with quality issues and or pathology present were included. These quality issues were considered in the same way as the AH images in Section 3.1.1.1. As far as possible the criteria for low quality status matched those of the Hoover et al. (2000) study. The distribution of these images can be appreciated from Table 3.2. The images were captured as Red Green Blue triplets, using 8 bits per plane. Only the green plane was used as the sole source of data because many of the images had a saturated red band. This saturation was due to the imaging process, it
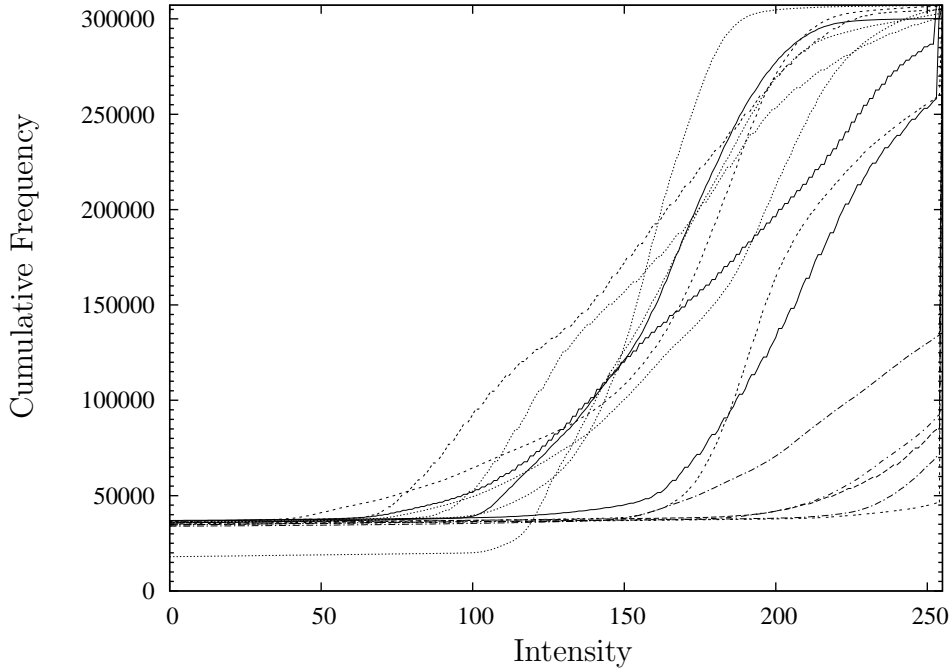
Figure 3.1: Cumulative frequency distribution of the red plane of the Royal Bournemouth Hospital image data set, illustrating the poor quality of this band and the reason it was rejected for processing

is illustrated by the cumulative frequency graph of figure 3.1. In the lower intensity range (0–100) the cumulative frequency is level at approximately 35k pixels, these picture elements belong to the dark perimeter or orientation mask around the image. From the graph it can be seen that the intensity range is not well used, and some images are almost totally devoid of non-saturated pixels. The single line outlying the group corresponds to an image with an unusually small orientation mask, as can be compared in figure 3.2.

3.1.2.1 Method: Labelling of the gold standard reference images

The ground truth was created by marking the vessels on the digital images using photo image editing software. The image was displayed on screen

<div align="center">a   b</div>
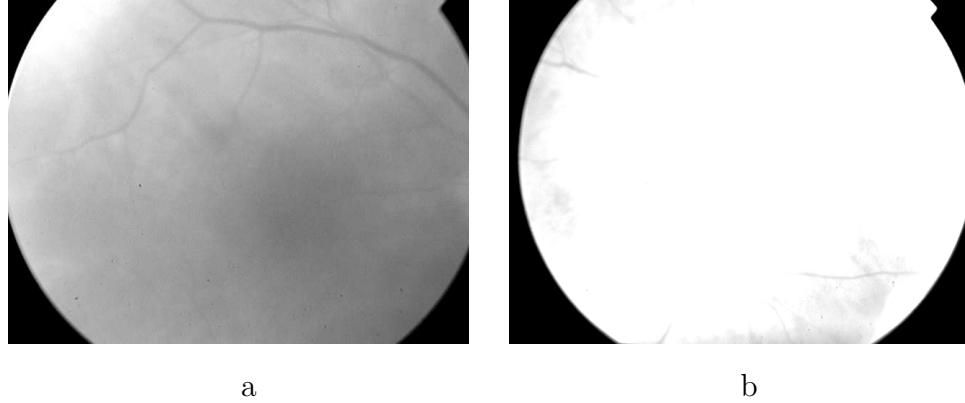
Figure 3.2: The red plane of images 5 (a) and 8 (b) of the Royal Bournemouth Hospital Data set; showing saturation of red sensor and different size of orientation mask.

overlaid with a transparent layer on which the marked image was created. This overlaid layer, or overlay had an alpha component so that the intensity of the retinal image modulates the intensity of the markings made on the transparent layer. In this way the markings did not obscure the vessels. It was felt that this feedback aided the accuracy of the markings. The exact methods employed are now given.

All the selected retinal images were labelled by the author using Adobe® Photoshop® . The 'raw' green plane was displayed in grey scale and an overlay-layer was added to each image on which the vessels were outlined using the pencil tool. Square brushes of one, four and nine pixels areas were used to mark the vessels within a range of magnifications (300%–1200%) depending on operator preference. The opacity of the overlay was set to 100%, i.e. fully transparent. The blend mode of the overlay was set to multiply and used the following values for the 'blend if grey' option: 0– 245. The underlying layer 'blend if grey' option was set to 0–249. The color used for marking was red. A sample of this labelling can be seen in Figures 3.3
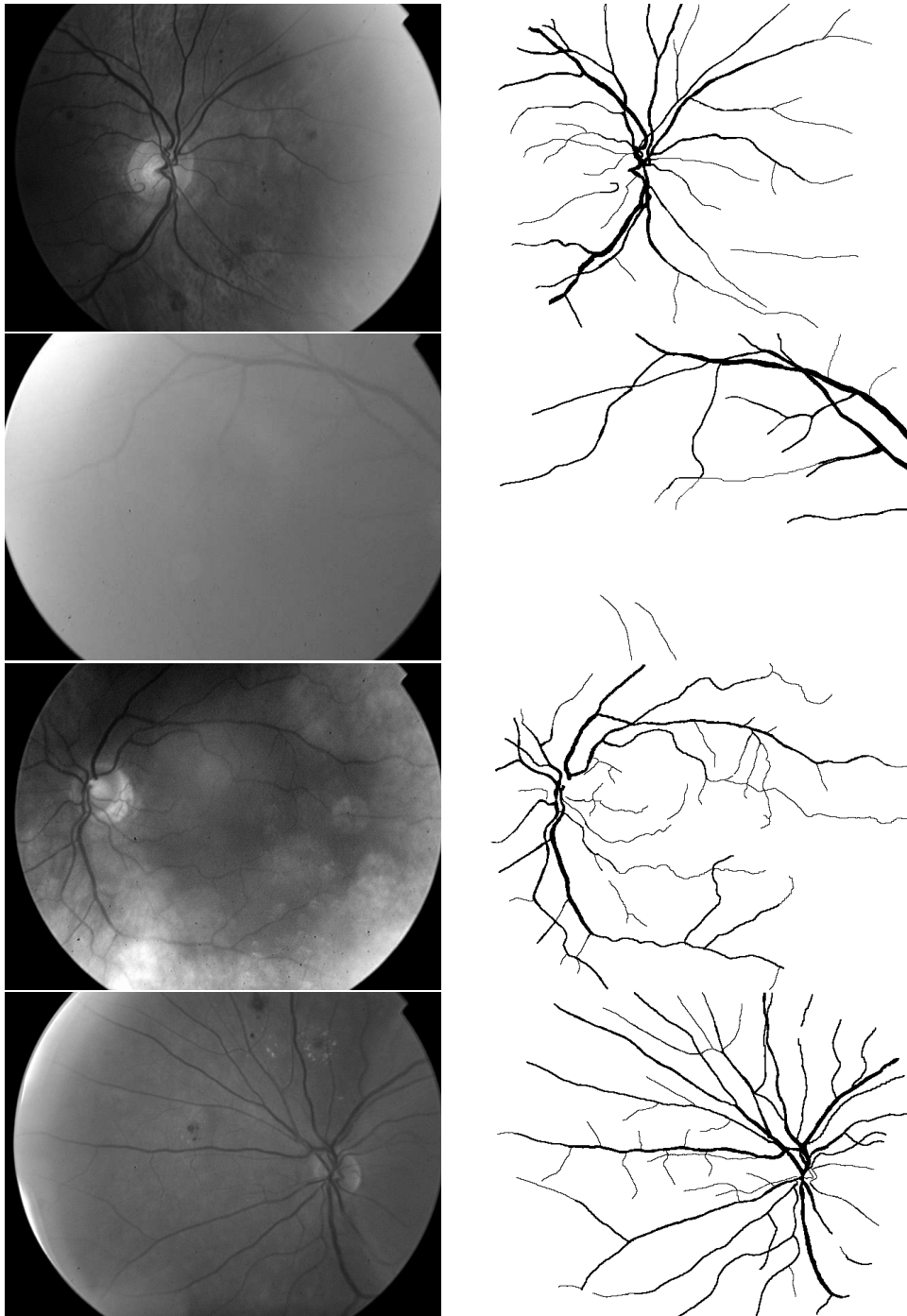
Figure 3.3: Green plane of Images 2, 5, 13 and 14 and their corresponding labelled ground truths on right

3.1.3 Introduction of methods and algorithm

The retinal vessel network was segmented using the green plane of the retinal images. A series of operations on this monochrome image produced bi-valued or binary images indicating the final vessel non-vessel status of each picture element. The results of this processing were prospectively evaluated by presenting images for processing which had not been used during the development of the algorithm. The Khoros software package[1] was used to implement the image processing tasks. Results are presented illustrating the effect of the inclusion or exclusion of the orientation mask. Performance of the algorithm on images with and without pathology and or quality-issues are also presented. The author used morphological operations at all stages of the processing, which can be broken down into three phases.

1. Preprocessing to remove the slower changes in intensity.

2. Identification of the vessels

3. Post processing removing the spurious pixels

Preprocessing was designed to remove the variations in illumination due to the curvature of the retinal surface and the crescent shaped reflections from the anterior ocular anatomy. Once illumination interferences are removed the remaining background regions are identified, for deletion, by their intensity and area. The area is calculated under a connectivity constraint, i.e. pixels must be connected to each other via a chain of neighbours. Areas of a given size and below threshold intensity are eliminated. The resulting image allows a single threshold to act as a decision boundary to dichotomize the remaining pixels into vessel and non-vessel. In the post processing step the binary image produced by the threshold is eroded

---

[1]Available : http://www.khoral.com/ Accessed 20th Sept. 2001

and dilated. The classification accuracy of these steps is then measured by comparison to the ground truth labelled images.

For a single retinal image, the algorithm described thus far has a fixed classification behavior. The parameters controlling it are static. However, many tradeoffs exist; one of these tradeoffs is achieved by lowering the threshold parameter so that more candidate pixels are included for area and intensity based discrimination. Likewise, decreasing the area by which pixels are rejected for post processing has the effect of relaxing the criteria by which they are accepted as vessels. In effect the morphological operations trade the 'ability' in detecting the vessels against the 'ability' to detect non-vessels.

This tradeoff is called the Receiver Operating Curve. ROCs stem from psychophysics, where there was a need for dis-associating the confidence or human decision making process from the underlying physics of the stimulus and sensory organ (Egan, 1975). In the context of the authors research ROCs negate, or account for, the selection of single parameter. The curves going a small way towards separating an algorithm from a particular choice of parameters and implementation. Apart from the benefits of providing a better characterisation of the algorithm, a further benefit is derived from being able to move the operating point to minimise missed vessels pixels at the expense of missed non-vessel pixels and vice versa.

To investigate this trade off in the two types of accuracy measurements the intensities at which the connected component parameter was evaluated is varied. This tightened or relaxed the criteria by which pixels were classified as non-vessels. The evaluation used two independent image data sets and their corresponding ground truths. These images were not used in development of the algorithm and were reserved to evaluate generalization ability.

3.1.3.1 Morphological image processing used in the study.

The methods used are based upon the application of set theory on binary (or bi-valued) and greyscale images and are well documented in image processing textbooks (Schalkoff, 1989; Gonzalez and Woods, 2002). Binary images may be referred to as a set of points $\boldsymbol{A}$, so that an individual element in the set $\boldsymbol{A}$ is indexed by the coordinates $\boldsymbol{x}$ and $\boldsymbol{y}$ by : $\boldsymbol{A}(\boldsymbol{x}, \boldsymbol{y})$. To distinguish the 8 bit greyscale images used in this study they are addressed by the lower case symbols such as $\boldsymbol{a}(\boldsymbol{x}, \boldsymbol{y})$ so that :

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{x}, \boldsymbol{y}) &\in \{\boldsymbol{0}, \boldsymbol{1}\} \quad \text{for binary images} \\ \boldsymbol{a}(\boldsymbol{x}, \boldsymbol{y}) &\in \{\boldsymbol{0}, \boldsymbol{1} \ldots \boldsymbol{N} - \boldsymbol{1}\} \quad \text{for greyscale images} \end{aligned} \tag{3.1}$$

In this study all images had $\boldsymbol{N} = \boldsymbol{256}$.

Morphological operations use the shape and the connectivity of regions of pixels to grow or shrink them. These operations use a second image, termed the structuring element to sweep in a raster style over the image being processed. Its action expands or contracts any underlying shapes that are present. This expansion and contraction corresponds to dilation and erosion; the two basic or primitive operators in morphology.

Erosion of image or set $\boldsymbol{A}$ by a structuring element $\boldsymbol{B}$ is denoted by the symbol $\ominus$. Erosion for binary images may be expressed by:

$$\boldsymbol{A} \ominus \boldsymbol{B} = \{\boldsymbol{x} | (\boldsymbol{B})_{\boldsymbol{x}} \subseteq \boldsymbol{A}\} \tag{3.2}$$

The operation equals the set of elements $\boldsymbol{x}$ for which all translations of $\boldsymbol{B}$ by $\boldsymbol{x}$, denoted by $(\boldsymbol{B})_{\boldsymbol{x}}$, are inside $\boldsymbol{A}$. The set of points occupied by the origin of $\boldsymbol{B}$ (show in Figure 3.4 as the light grey square) is the resulting set or image. Figure 3.4 illustrates this operation.

Dilation is defined is a similar fashion to erosion. The dilated image results from the union of the image and structuring element for all possible translations of the structuring element $\boldsymbol{B}$ for which the origin is still within $\boldsymbol{A}$. The operation is denoted by the symbol $\oplus$ and for the binary case with
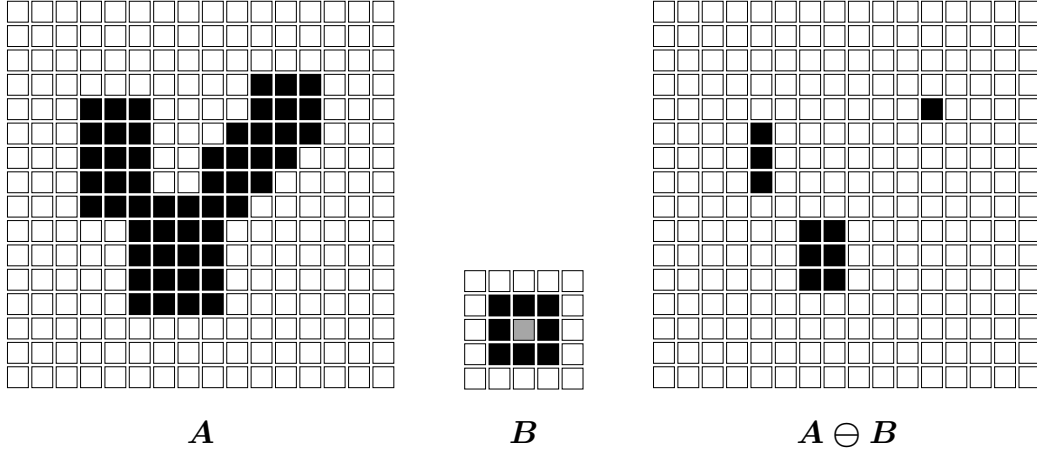
Figure 3.4: Example set or image (left), structuring element (centre, lighter pixel in structuring element is considered the origin) and the results of the erosion (right).
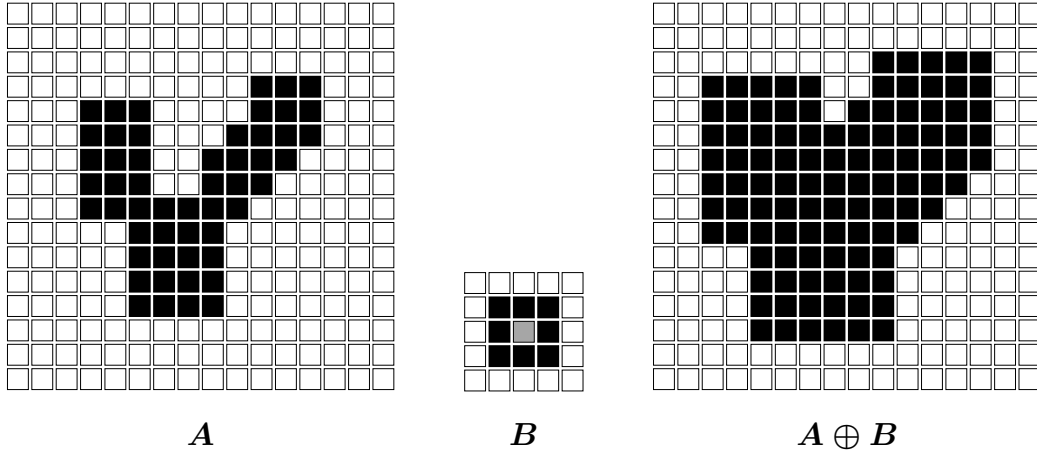
Figure 3.5: Dilation of set or image $\boldsymbol{A}$ by structuring element $\boldsymbol{B}$

a symmetrical structuring element the operation may be given by :

$$A \oplus B = \{x | (B)_x \cap A\} \tag{3.3}$$

The dilation and erosion operators can also be extended to greyscale or monochrome images. The min() and max() operators return the maximal or minimal value of their argument, they are used to obtain a single result

for each position of the structuring element. For dilation this may be expressed as :

$$
\begin{aligned}
(a \oplus b)(s,t) \;=\; \max\{ \quad & a(s-x, t-y) + b(x,y)| \\
& (s-x), (t-y) \in D_a \\
& (x,y) \in D_b \ \}
\end{aligned}
\tag{3.4}
$$

where positions of the structuring element are constrained by the requirements that coordinate pairs $(s-x), (t-y)$ and (x,y) fall within the original image and the structuring element. In other words $D_{\{a,b\}}$ stipulates that the structuring element b and original image a overlap. The operation is similar for erosion:

$$
\begin{aligned}
(a \ominus b)(s,t) \;=\; \min\{ \quad & a(s+x, t+y) - b(x,y) \ | \\
& (s+x), (t+y) \in D_a \\
& (x,y) \in D_b \ \}
\end{aligned}
\tag{3.5}
$$

A graphical representation of grey scale erosion and dilation in the two dimensional case is presented in Figure 3.6. Here the solid line represents a section, profile, row or column traversing the image or structuring element. As these images or functions slide over each other, the maximum or minimum of their supersition is kept as the resulting image. For both greyscale and binary images dilation and erosion operations are usually applied in sequence, either the dilation is applied first the result of which is subsequently eroded. This sequence is called 'closing' and is denoted by the ● symbol. Thus the closing of A by structuring element B is expressed as follows:

$$
A \bullet B = (A \oplus B) \ominus B
\tag{3.6}
$$

in the binary case. Opening, in which the erosion and dilation operations are reversed, may be defined as follows :

$$
A \circ B = (A \ominus B) \oplus B
\tag{3.7}
$$
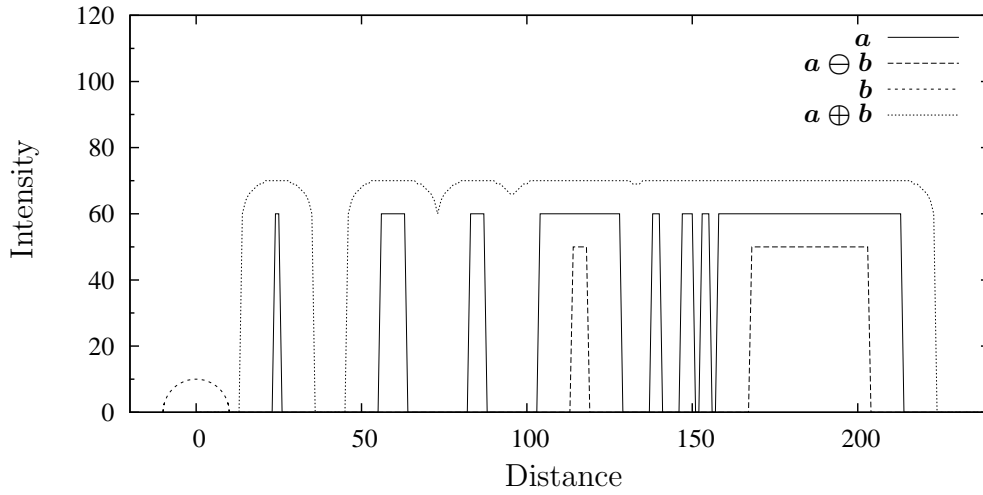
Figure 3.6: Grey scale erosion and dilation of a using a circular structuring element b on a sample 2D profile.

In grey scale images these become

$$a \bullet b = (a \oplus b) \ominus b \tag{3.8}$$

and

$$a \circ b = (a \ominus b) \oplus b \tag{3.9}$$

The results of applying opening and closing to the binary image used in Figures 3.4 and 3.5 is illustrated in figure 3.7. A visual representation of opening and closing for grey scale images is provided in Figure 3.8.

The binary dilation operator is now iteratively extended to extract groups of connected pixels. In a square grid, a pixel is said to be 8-way connected (N_8) to all of its eight neighbours. The centre pixel in the structuring element of Figure 3.5 is a good example. In a similar fashion a 4-way (N_4) relationship can be defined, with the North, East, South and

$$A \qquad\qquad A \circ B \qquad\qquad A \bullet B$$
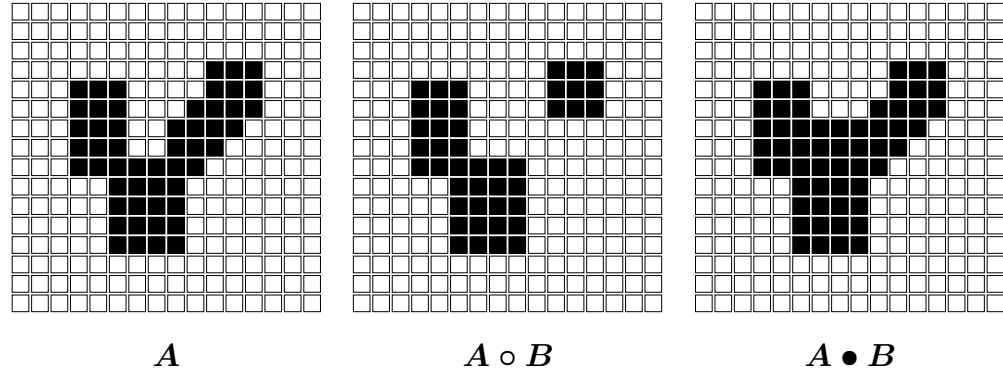
Figure 3.7: Results of opening and closing image A using the same structuring element (B) used in the erosion and dilation of Figures 3.4 and 3.5

West pixels connected. A connected component is simply a group of these connected pixels. In terms of morphology the relationship maybe defined so that:

$$A_k = (A_{k-1} \oplus B) \cap A \qquad k = 1, 2, 3, \ldots \qquad (3.10)$$

this is an iterated equation, with k being the index to the iteration. A pixel inside the connected component is used as a starting point i.e. $p \in A$, and so long as the structuring element used is symmetrical the connected component is invariant to the selection of this seed point (provided $p \in A$). The structuring element B defines the type of connectivity that is used, N_8 if it is shaped as in figures 3.4 or 3.5. Once connected components have been identified the area occupied by the connected component's pixels is given by the number of its members, or the cardinality. This is expressed as:

$$\mathrm{Card}(A) \qquad\qquad (3.11)$$

The two connected components of $A \circ B$ in Figure 3.7 are shown in Figure 3.9. So that $\mathrm{Card}(A_a) = 9$ and $\mathrm{Card}(A_b) = 34$.
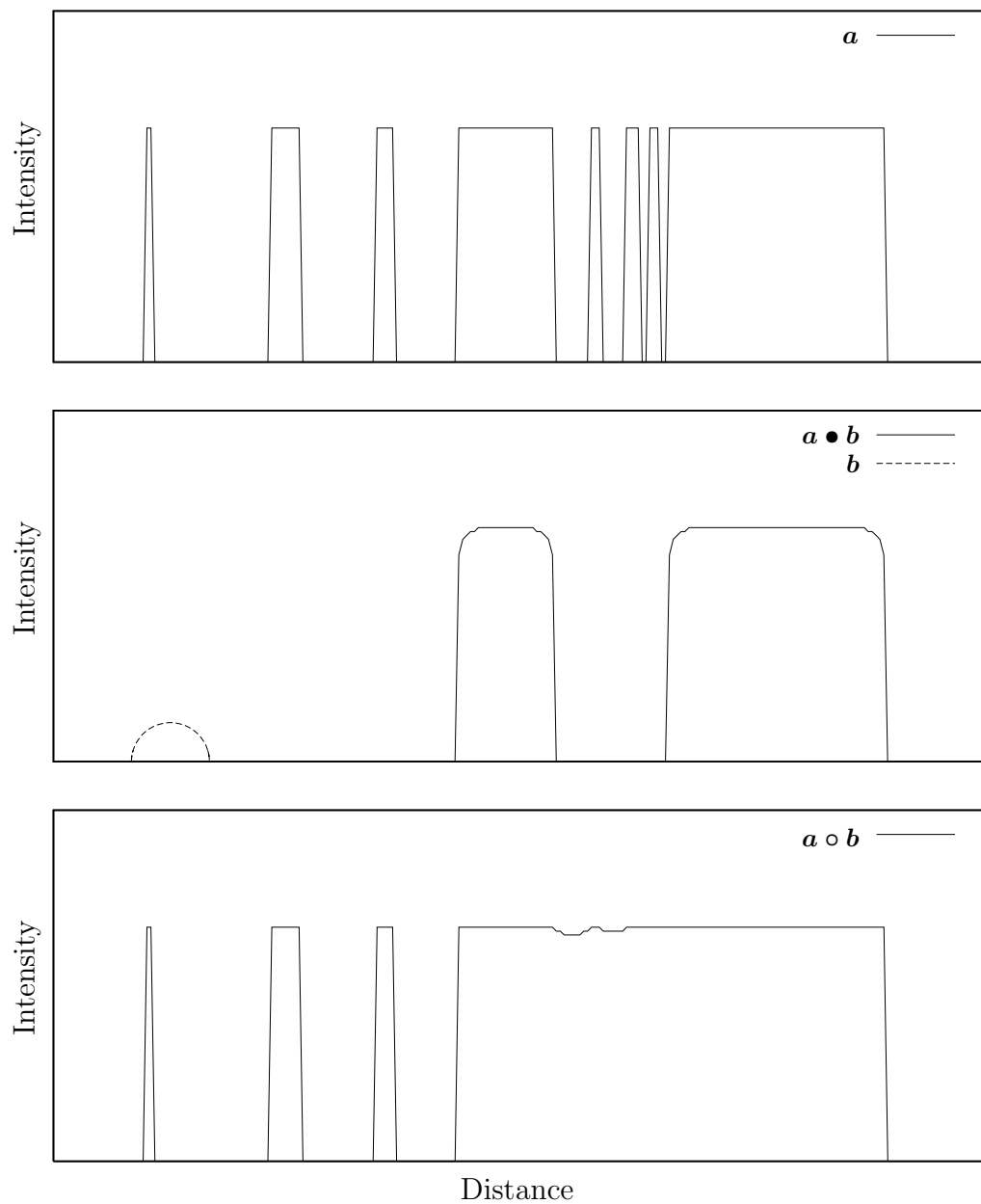
Figure 3.8: Effects of opening and closing on the sample profile. Note the elimination of the smaller 'valleys' in the closing operation.
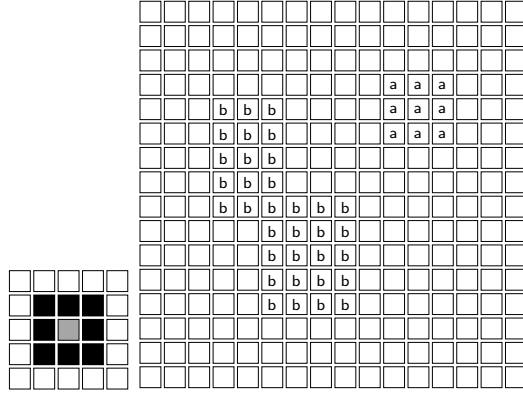
Figure 3.9: Two N_8 way connected components and associated structuring element.

3.1.4   Summary of morphological operations

The previous sections on mathematical morphology for image processing described all but a few of the operations used in the segmentation of the vascular network. This section shall give details on the order of operations, structuring elements and image masks that were used.

3.2   Methods

There are three main steps before the assessment of accuracy is made. These steps correspond to: an initial background subtraction via a grey scale opening, a threshold and closing and thirdly a connected component analysis. The second step is cycled over a range of thresholds. These methods are at first presented using a single image. A description of how they are iterated over entire data sets is given in Subsection 3.2.3. A reference to the symbols and their description is given in table 3.3.

Symbol	Description
Images (binary)	
AH	Hoover et al. (2000) images
$\boldsymbol{B}$	Three by three structuring element presented in Figure 3.5
$\boldsymbol{C}$	Thresholded image $\boldsymbol{c}$
$\boldsymbol{D}$	Image from intermediate state of processing, closed version of $\boldsymbol{C}$
$\boldsymbol{E}$	Labelled connected component image
$\boldsymbol{F}$	Image $\boldsymbol{E}$ with components of area $\boldsymbol{n}$ removed
$\boldsymbol{G}$	Final processed image
$\boldsymbol{M}$	Image masks used for perifery elimination
$\boldsymbol{R}$	Reference images, human labelled.
RBH	Images of this study
$\boldsymbol{TP}$	True Positive image
$\boldsymbol{TN}$	True Negative image
$\boldsymbol{FP}$	False Positive image
$\boldsymbol{FN}$	False Negative image
Images (real)	
a	Original unprocessed image
b	Spherical structuring element
c	Background subtracted image
Parameters	
$\boldsymbol{i}$	image identifier $\{\boldsymbol{0, 1 \ldots 14}\vert\mathrm{rbh}\}, \{\boldsymbol{0, 1 \ldots 19}\vert\mathrm{ah}\}$
$\boldsymbol{k}$	iteration counter of connected component analysis.
$\boldsymbol{n}$	area of connected component
$\boldsymbol{r}$	radius of hemispherical structuring element $\boldsymbol{b}$
rbh	Royal Bournemouth Hospital Images
$\boldsymbol{s}_{\mathrm{TPF}}^{\boldsymbol{t}}$	Standard Deviation of $\mathrm{TPF}_{\boldsymbol{t}}$
$\boldsymbol{s}_{\mathrm{FPF}}^{\boldsymbol{t}}$	Standard Deviation of $\mathrm{FPF}_{\boldsymbol{t}}$.
$\boldsymbol{t}$	Threshold level applied to image $\boldsymbol{c}$
$\boldsymbol{x, y}$	Index variables into the images
TPF	True Positive Fraction or sensitivity
TNF	True Negative Fraction or specificity
FPF	False Positive Fraction
$\mathrm{T\overline{P}F}_{\boldsymbol{t}}$	Mean True Positive Fraction or sensitivity
$\mathrm{F\overline{P}F}_{\boldsymbol{t}}$	Mean False Positive Fraction or false hit rate

Table 3.3: Symbols used in the morphological algorithm and accuracy assessment

3.2.1 Grey scale morphological opening : Rolling ball background subtraction

The green plane of the retinal image was equalized or shade corrected using an erosion followed by a dilation, i.e. a morphological opening. In this case the structuring element used was spherical, and the result was subtracted from the original image. This opening followed by a subtraction is sometimes referred to as 'a rolling ball background subtraction' (Gonzalez and Woods, 2002). The structuring element b may be represented using :

$$b(x, y) = \sqrt{r^2 - x^2 - y^2}, \quad -\tfrac{r}{2} < x < \tfrac{r}{2},$$
$$-\tfrac{r}{2} < y < \tfrac{r}{2} \tag{3.12}$$

where r is the radius of the spherical structuring element. The opened image was then subtracted from the original :

$$c = (a \bullet b) - a \tag{3.13}$$

The effects of these operations are illustrated by processing the image in Figure 3.10 and plotting a section or two dimensional profile of each step.

3.2.2 Area and intensity-range constrained particle analysis

This step is aimed at further identification and removal of the non-vessel background. Operating on background-subtracted image c, a global threshold t is applied :

$$C(x, y) \;=\; \begin{cases} 1, & c(x, y) < t \\ 0, & \text{otherwise} \end{cases} \tag{3.14}$$

The thresholded image C was then dilated twice and eroded or :

$$D = (C \oplus B) \circ B \tag{3.15}$$
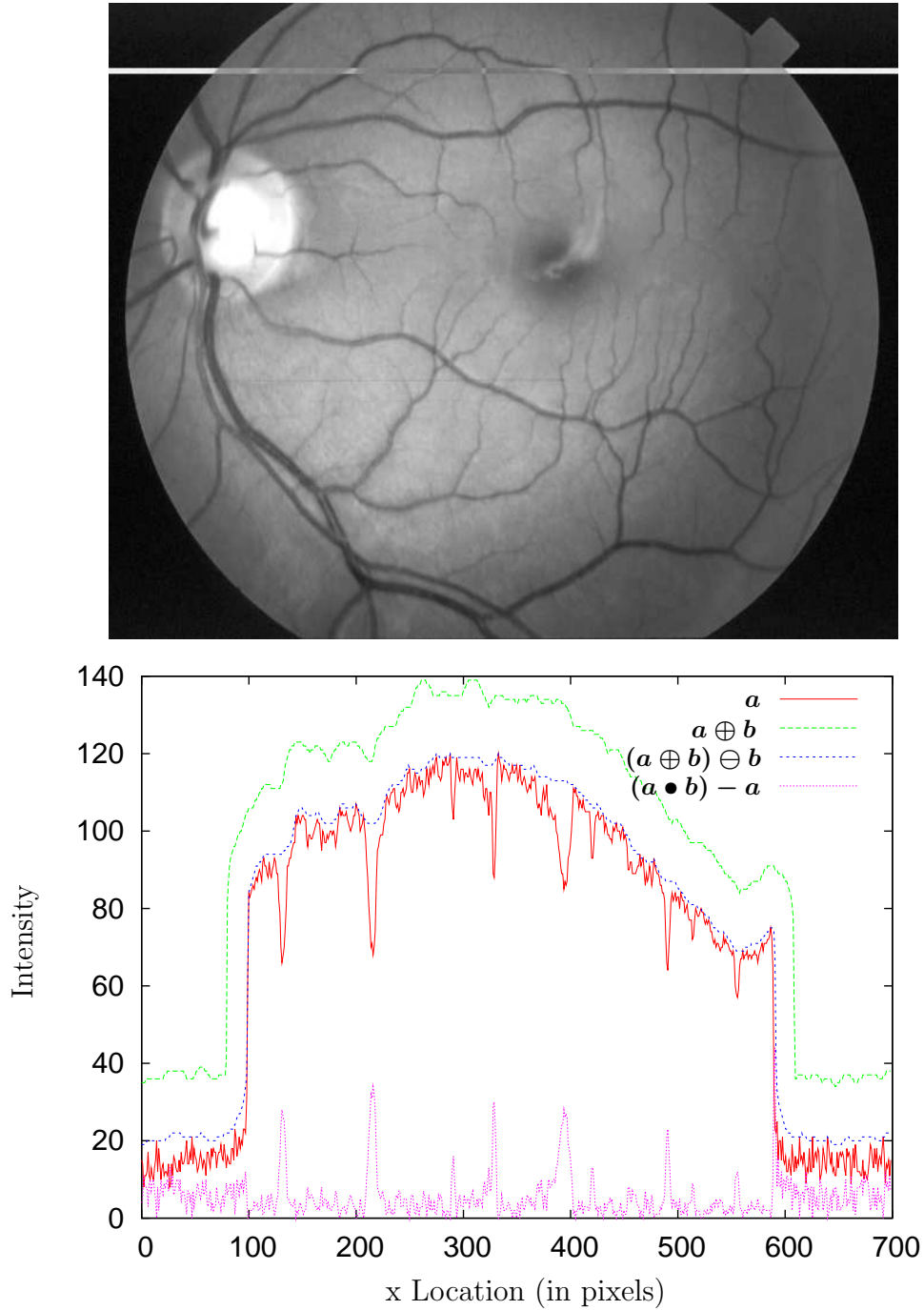
Figure 3.10: Image 17 from the Hoover et al. (2000) data with sample plots taken from horizontal line one tenth from top. It illustrates the effect of the dilation followed by the erosion, or opening, and how the opening subtracted from the original image produces the image profile adjacent to the ordinate axis.

The resulting image was analysed for N_8 connected components. These connected areas were set to zero if they containing less than n pixels. Thus for each picture element in D:

$$E_k = (E_{k-1} \oplus B) \cap D \qquad (3.16)$$

that is connected to less than n other picture elements is set to zero. I.e:

$$F = \begin{cases} 0, & \text{Card}(E_k) < n \\ 1 - C, & \text{otherwise} \end{cases} \qquad (3.17)$$

In this study n was chosen to be 15 pixels in area. This parameter was chosen by interactively a sweeping threshold *bands* over the low (0 to 30) intensity pixels. This highlighted a type of shot, or speckled, noise, that was characterised by their small areas. These noise features are visible as the medium height peaks of the '$(a \bullet b) - a$' trace in Figure 3.10

Once the smaller particles are removed the image orientation place holder is removed by masking it with image M_{rbh} for the Royal Bournemouth Hospital data, and image M_{ah} for the Hoover et al. (2000) data. Both masks are shown in Figure 3.11. The black areas of the mask are zeroed, whilst the white central region is kept. That is:

$$G = F \cap M_{\{\text{rbh,ah}\}} \qquad (3.18)$$

The image G is the binary image representing the classification into vessel and non-vessel classes. This final image is evaluated against the manually labelled images detailed in Section 3.1.2.1. These human labelled images R provide a reference from which two types of accuracies will be derived. The two accuracies give the fidelity of the morphological segmentation process for the identification of vessel and non-vessel picture elements. These two descriptors arise out of the four classification possibilities resulting from the state of the predictions given by image G and labelling from R.

$$G$$

		+ve	-ve
	True	TP	FN
R	False	FP	TN

An alternative definition of the above table is given by:

$$
\begin{aligned}
\boldsymbol{TP} &= \boldsymbol{G} \cap \boldsymbol{R} \\
\boldsymbol{TN} &= \boldsymbol{G^c} \cap \boldsymbol{R^c} \\
\boldsymbol{FP} &= \boldsymbol{G} \cap \boldsymbol{R^c} \\
\boldsymbol{FN} &= \boldsymbol{G^c} \cap \boldsymbol{R}
\end{aligned}
\tag{3.19}
$$

$\boldsymbol{A^c}$ indicates $\boldsymbol{A}$'s complement or $\boldsymbol{A^c} = \{\boldsymbol{x}, \boldsymbol{x} \notin \boldsymbol{A}\}$. With the cardinality of TP, TN, FP and FN the sensitivity or True Positive Fraction is defined as

$$
\mathrm{TPF} = \frac{\mathrm{Card}(\boldsymbol{TP})}{\mathrm{Card}(\boldsymbol{TP} \cup \boldsymbol{FN})}
\tag{3.20}
$$

similarly the True Negative Fraction or specificity is defined as:

$$
\mathrm{TNF} = \frac{\mathrm{Card}(\boldsymbol{TN})}{\mathrm{Card}(\boldsymbol{TN} \cup \boldsymbol{FP})}
\tag{3.21}
$$

these two definitions give the two fractional accuracies of the classification process. Ideally $\boldsymbol{FP} = \boldsymbol{FN} = \emptyset$ for a sensitivity and specificity of 100%. However perfect operation is not possible, and usually a trade-off between the accuracies is possible by varying the criteria of the detection. In this study, that criterion was the threshold parameter $\boldsymbol{t}$. How this parameter was varied to produce the ROC curves is described in the next section.

3.2.3   Iterated aspects of the method

The segmentation method of the previous section assumes all the parameters, such as $\boldsymbol{r}, \boldsymbol{n}$ and $\boldsymbol{t}$ are static and that all the operations are applied to a single image to produce a single TPF, TNF result-pair. However, this study has two datasets with many images in each. Subsets have also been created from each dataset (detailed in Tables 3.2 and 3.1) This section will

$$\boldsymbol{M}_{\text{ah}} \qquad\qquad\qquad\qquad \boldsymbol{M}_{\text{rbh}}$$
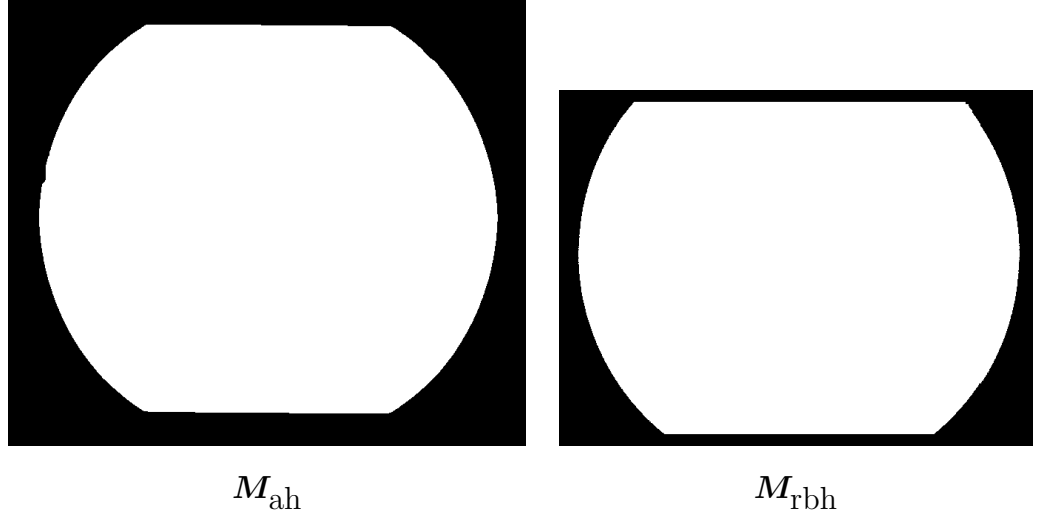
Figure 3.11: The two masks used to exclude image perifery. $\boldsymbol{M}_{\text{ah}}$ from Hoover et al. (2000)'s study excluded 34% of pixels. Like wise mask $\boldsymbol{M}_{\text{rbh}}$ excluded 22% in images from the Royal Bournemouth Hospital.

describe how aspects of the segmentation procedure are iterated upon an individual image, image subsets and image datasets. The section will also describe how these results are displayed, summarised and compared by means of ROC curves.

Starting with a single image from either data set, this is processed multiple times for a range of values of $\boldsymbol{t}$. That is; each image has a single rolling ball background subtraction after which it is thresholded, particle analysed and morphologically closed multiple times for many different values of $\boldsymbol{t}$. The range of intensity values chosen for $\boldsymbol{t}$, spans 0–100, in increments of one. The threshold operation of equation 3.14 now becomes:

$$\boldsymbol{C_t}(\boldsymbol{x}, \boldsymbol{y}) \;=\; \begin{cases} \boldsymbol{1}, & \boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y}) < \boldsymbol{t}\,, \;\; \boldsymbol{0} \leqslant \boldsymbol{t} \leqslant \boldsymbol{100}, \; \boldsymbol{t} \in \mathcal{N} \\ \boldsymbol{0}, & \text{otherwise} \end{cases} \qquad (3.22)$$

A series of images, $\boldsymbol{C_t}$, is thus obtained. These instances at various thresholds can then be processed in further as described in equations 3.16 through to 3.21.

The data sets contain more than one image, so that a combination of a (the original image) and t are no longer sufficient to identify G (the resulting output image). Another subscript, in this case i is used to identify the images. The distinction between the RBH and AH datasets is left to the context of the results. The sensitivity or TPF and specificity TNF are thus given by $\text{TPF}_{i,t}$ and $\text{TNF}_{i,t}$ to indicate the image and threshold at which they were calculated. A table giving the results over a threshold range for every image in a dataset can then be constructed. Using a pair of columns $\text{TPF}_{i,t}$ and $\text{TNF}_{i,t}$ for each image i and a row for each threshold t. These data are voluminous and visual representation can provide a better overview of the data.

3.2.4   Visualisation of the results

When each retinal images is classified there are many sensitivity and specificity pairs that are generated over the range of threshold values. When these points, $\text{TPF}_{i,t}$, $\text{FPF}_{i,t}$, are sequentially linked the ROC curve for that images is created. Examples of these ROC curves for 5 images over $0 \leqslant t \leqslant 100$ are given in Figure 3.12. Each curve arching from 0,0 to 1,1 and getting close to 0,1 is formed by the locus of points that represent the operating tradeoffs between true positive detection and true negative detections for one image.

The ROCs in Figure 3.12 also demonstrate the worst and best vessel identification accuracies and associated false positive rates for that set or collection of images. These are the most exterior traces, or the ones that are closest to the 0,1 and 1,0 points. In Figure 3.12 these maximum and minimum performance delimeters belong to classification accuracies obtained from Image 10 and 13. In results presented later in this chapter these delimiters of performance will be used to illustrate how performances are bounded for a particular image set.
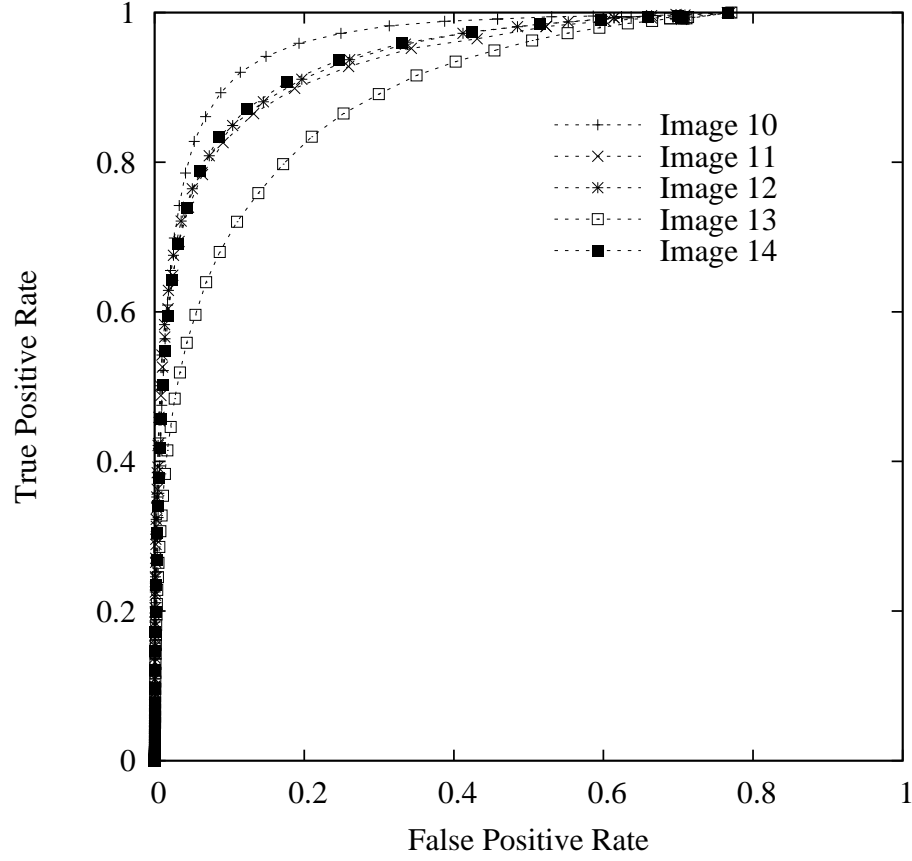
Figure 3.12: Range of operating points over $\mathbf{0} \;\leqslant\; \boldsymbol{t} \;\leqslant\; \mathbf{100}$ for the five images (10–14) used in the verification of the morphological segementation of the RBH dataset.

Presenting the maximum and minimum of the classification tradeoffs does not communicate the distribution of operating points for a particular value of $\boldsymbol{t}$. This distribution is illustrated by plotting the same ROCs in Figure 3.12 as a scattergram where the points are in identical positions and are keyed not by their image number, but by their threshold value $\boldsymbol{t}$. This alternate visualisation is shown in Figure 3.13 where a few values of $\boldsymbol{t}$ have been selected and the performance point for each image identified.

The scattergrams presented thus far are adequate to present results for
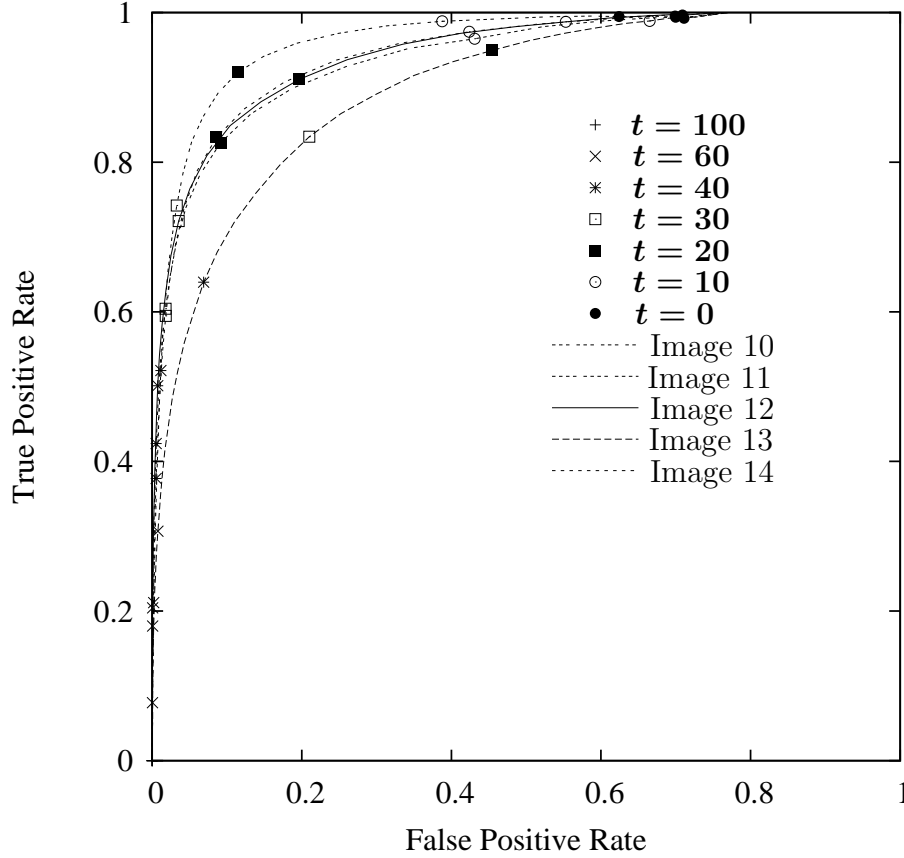
Figure 3.13: Selected operating points keyed by threshold value t for images (10–14) of the RBH dataset. This figure shows how for a set of five images at a single threshold values how distributions of operating points arise. Thus for each t value we have a range, mean and distribution of TFP and FPF values.

a handful of images or for a few t values. However, to oversee the distribution of TPF, FPF pairs for all threshold values, a more concise presentation is needed. This is achieved using the mean and standard deviation, defined in the usual manner, so that:

$$\overline{\text{TPF}}_t = \frac{1}{n} \sum_i \text{TPF}_{t,i} \qquad (3.23)$$

and

$$s^t_{\mathrm{TPF}} = \sqrt{\frac{1}{n}\sum_i (\overline{\mathrm{TPF}}_t - \mathrm{TPF}_{t,i})^2} \qquad (3.24)$$

Substituting $\mathrm{TPF}_{t,i}$ for $1\text{-}\mathrm{TNF}_{t,i}$ produces the mean and standard deviation for the False Positive Fraction. An example of how the mean specificity and sensitivity, and their respective standard-deviations, vary with t is shown for the AH data in Figure 3.14.

The standard deviation is not the best measure to describe the spread of a quantity tending towards a limit. The region of the TPF, FPF graphs where this statistic is used retains 10 to 20% of its potential range.
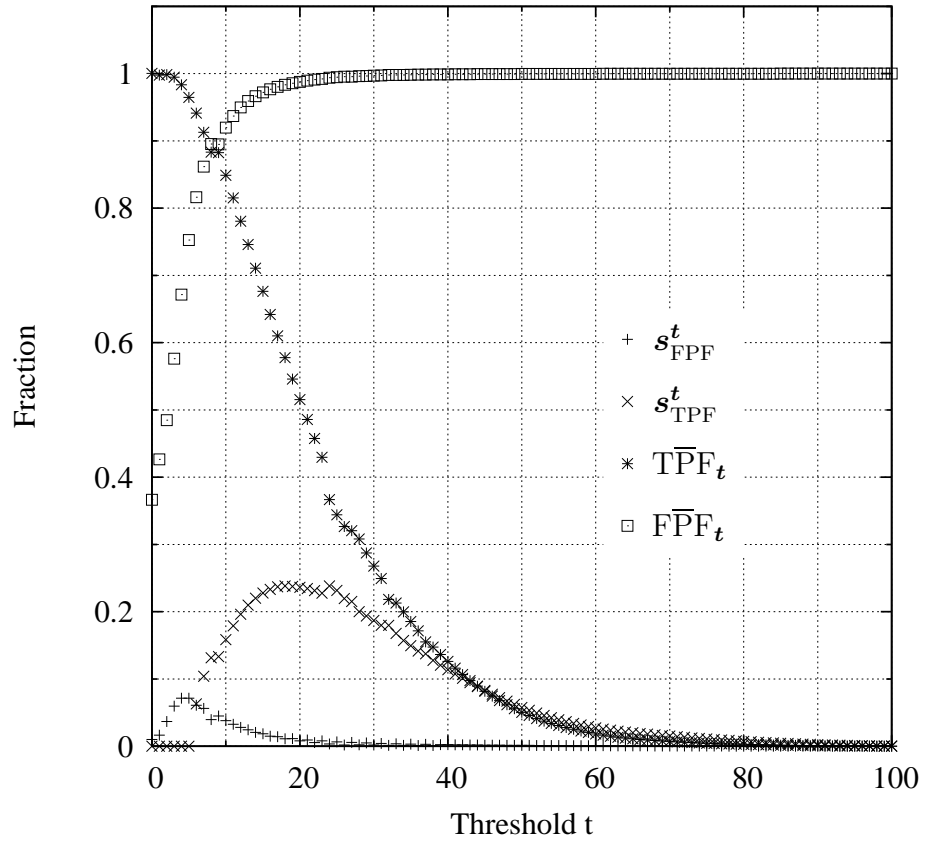
Figure 3.14: Sample plot of mean and standard deviation of sensitivity and false positive fraction (1-specificity) over the range of thresholds used.

The more usual way of displaying the sensitivity and false positive rates in Figure 3.14 is via a ROC curve that plots the $\mathrm{T}\overline{\mathrm{PF}}_t$ against $\mathrm{F}\overline{\mathrm{PF}}_t$. These 'mean vs mean' ROCs enable a cursory comparison of segmentation performance across different subsets, datasets and methods of processing. An example of this type of plot is given in Figure 3.15, where the tradeoffs between the abilities to detect vessels and the ability to detect non-vessels is compared for the RBH development and verification data.

3.2.5 Method Summary

The morphological methods described are based on area or pixel by pixel operations. These operations are not new and have been found in standard texts a for over a decade (Schalkoff, 1989). These methods do however take the novel step of assessing the performance of these morphological methods. Segmentation ability of the algorithm is compared to human labelled ground-truth images a to produce more objective representations in the form of ROCs.

The method starts with background subtraction, which is somewhat analogous to high pass filtering. The thresholded background-subtracted image is then analysed for connected components which are assumed to represent the identified vessels.

The resulting binary image is compared to the binary image taken to be the ground truth. The comparison of the two images gives rise to the four possible pixel types. The fraction of correctly identified vessels out of all actual vessel pixels (TPF) and the fraction of detected non-vessel pixels(TNF) out of all actual non-vessel pixels are the two measurements taken from the classification process.

The sensitivity (TPF)and the specificity (TNF) are then measured over a range of a decision parameter t for both the local and AH datasets with and without the masking operation. Further image test sets arise out of
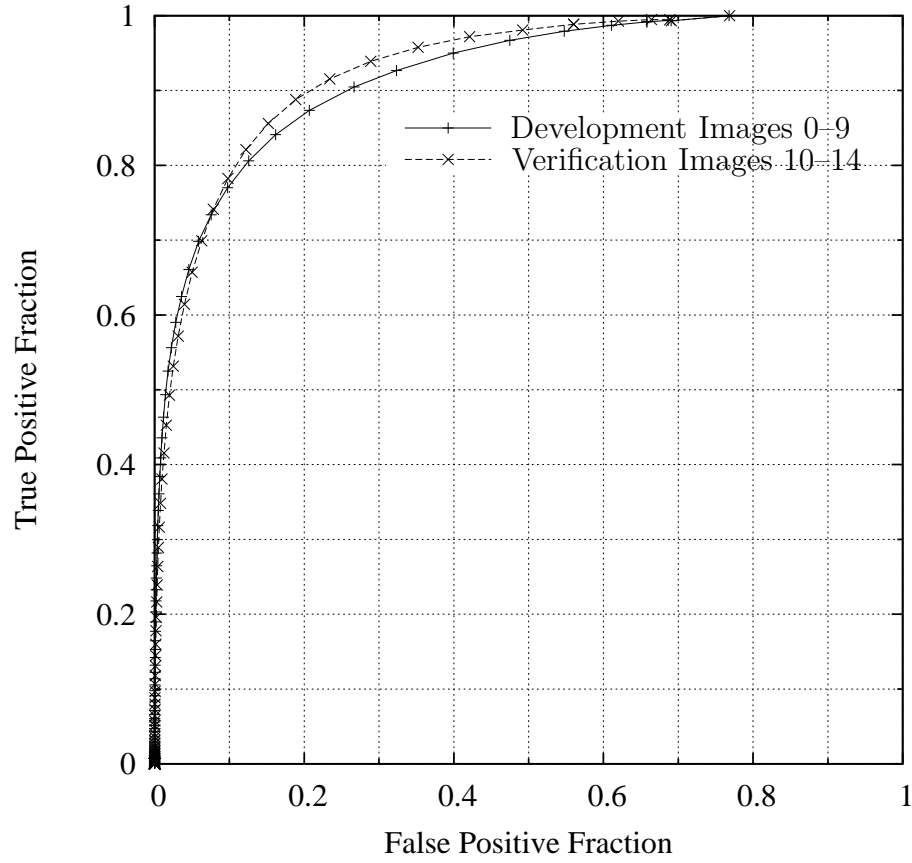
Figure 3.15: Example plot of mean sensitivity against mean false positive fraction ($\mathrm{T\overline{P}F}_t$, $\mathrm{F\overline{P}F}_t$) with $\mathbf{0} \leqslant \boldsymbol{t} \leqslant \mathbf{100}$ for the development and verification data. At low threshold values, i.e. towards the right of the graph, there are fewer false positives in the verification images than the development data.. The curves also cross, illustrating that a single point on the ROC curve is not sufficient to identify a curve among a family of curves. Although in this case it is very possibly due to a difference in the underlying datasets.

divisions or amalgamations of the data sets. Details on these subsets are presented in the next section.
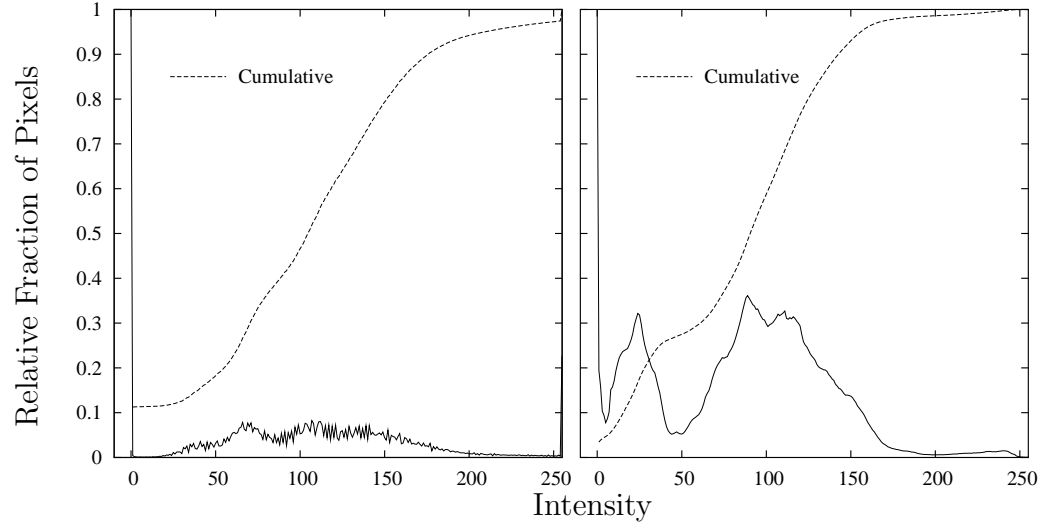
Figure 3.16: Normalised intensity histograms and cumulative pixel counts of all RBH (left) and all AH (right) data. The curves were normalised so that $(\boldsymbol{y}/\max(\boldsymbol{y}))$

3.3   Results

The results describing the image data and the outcomes of the methods presented via their operating characteristics will be presented in this section. Initially histograms will be presented that summarise the images from Tables 3.2 and 3.1. Then TPF:FPF tradeoffs for individual datasets and comparisons between them will be given. Actual instances or graphic representation of the algorithm will be given at the end of this section.

Only the green plane of the images was used, the distribution of pixel intensities for the AH and RBH datasets is now given. The distributions in Figure 3.16 represent all images, i.e. images 0–14 for the RBH data, and images 0–19 for the AH data. Note, frequencies and sums are remapped onto the 0–1 range because absolute counts are too disparate to make intra data(sub)set comparisons.

Histograms for the various subsets are now presented. First of all, the
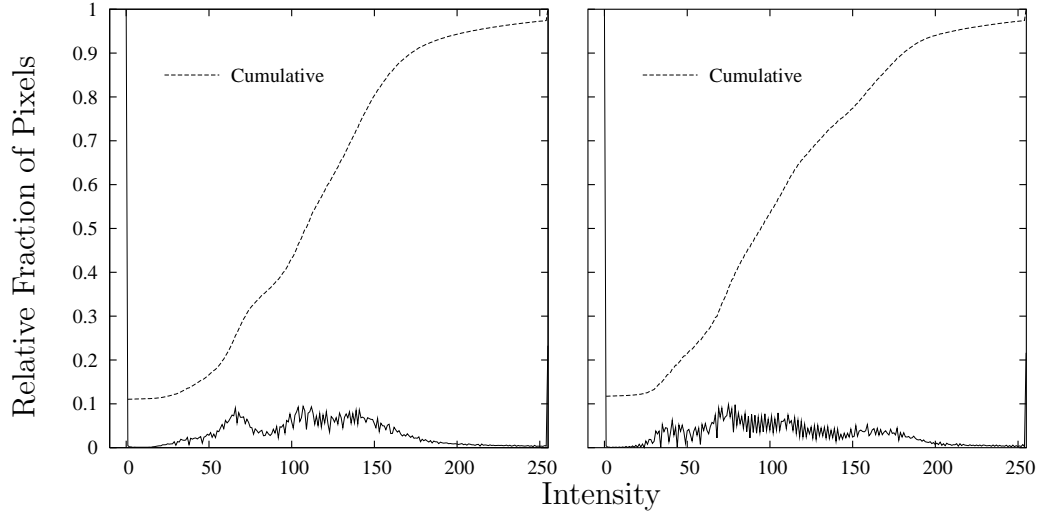
Figure 3.17: Intensity histograms of RBH development (left) and verification (right) data. Note these graphs have been normalised ($\boldsymbol{y}/\max(\boldsymbol{y})$) and that both the default intensity (i.e. not the cumulative traces) histograms have their maxima at 1, for zero intensity.

split between development and verification subsets in the RBH data is illustrated in Figure 3.17. Figure 3.18 graphs the intensity histograms and cumulative histograms for the AH dataset split into normal and pathological subsets.

The distribution of intensity values that arise when the pixels are keyed by their status (as given by the labelled images) are presented in Figure 3.19 for the RBH data and in Figure 3.20 for the AH data. RBH images had on average 5.5% of pixels ground truth labelled as vascular, where as the AH images had 7.6%.

The AH dataset and the RBH data set were processed twice, once with the masking operation described in Equation 3.18 and once without. These results, and the results of some of the possible subsets are presented in the next sections and are organised as follows:

Section 3.3.1 presents results for the locally labelled dataset from the
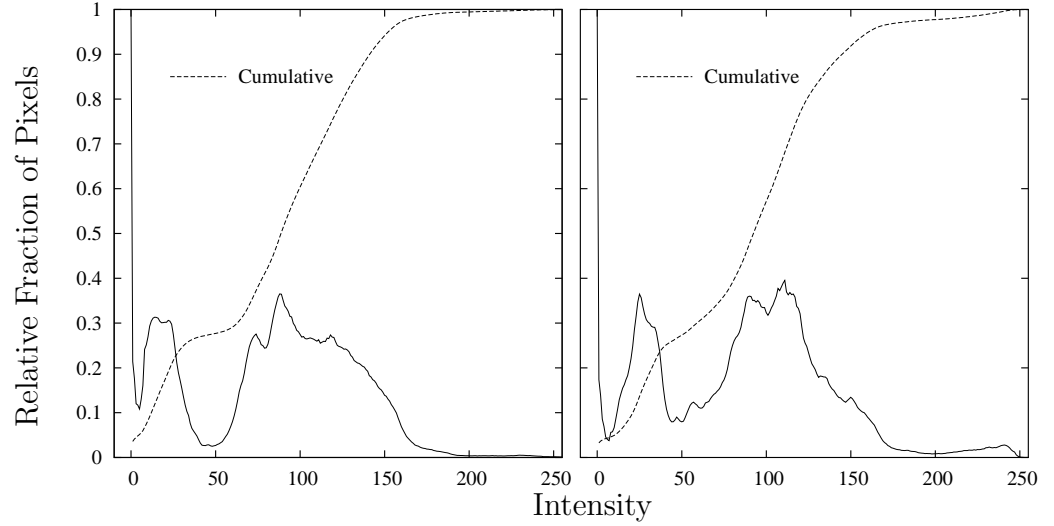
Figure 3.18: Intensity histograms showing the Normal (left) and pathological (right) AH data.

Royal Bournemouth Hospital.

Section 3.3.2 presents results for the Hoover et al. (2000) labelled dataset.

Section 3.3.3 compares results for the omission or inclusion of the masking operation of Equation 3.18.

Section 3.3.4 displays results of images containing pathology or not, with or without imaging quality considerations.

Section 3.3.5 details the classification accuracies obtained when image quality is a factor.

Section 3.3.6 compares the operation of the segmentation on the two main datasets.

Section 3.3.7 details the classification accuracies obtained when image quality and pathology are considered in combination
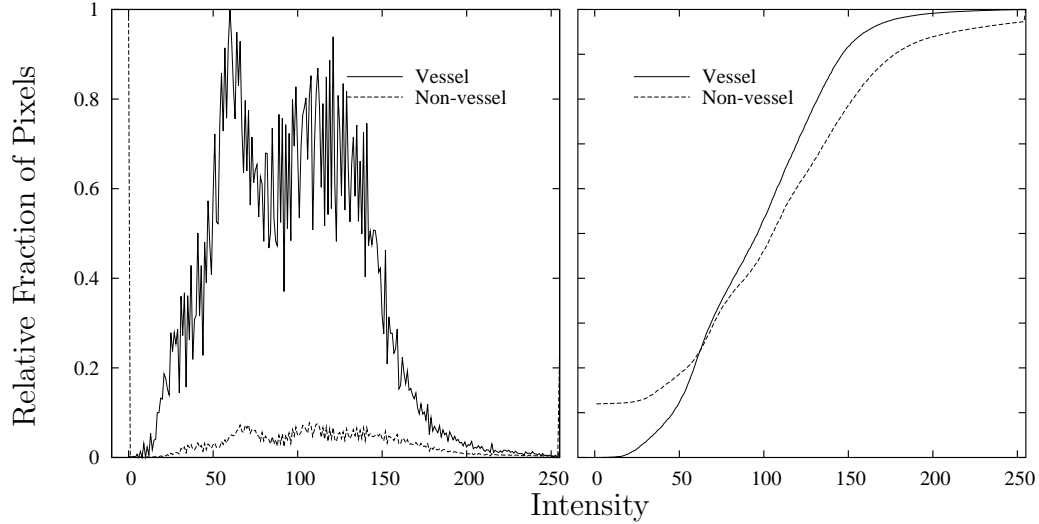
Figure 3.19: Intensity histograms of all RBH data showing pixel-frequencies keyed by their vascular status as given by the ground truth labeling. Erratic pixel frequencies were double checked (For the vessel trace, 1 = 3156 pixels. For the non-vessel trace 1= 516871 pixels) Cumulative intensities are given on the right.

Section 3.3.8 presents a sample of classified images to illustrate the operation of the algorithm.

All results were obtained using a rolling ball radius (r) of 15. This radius was used because vessels were seldom wider than 15 pixels and therefore the morphological opening would not eliminate these features. Examples of the output images are displayed at the end of the results section, prior to the conclusion.

3.3.1 Development and prospective performance validation results

The graphs of Figures 3.21 describe final results from the original images used in the development of the algorithm. These images were used to 'visu-
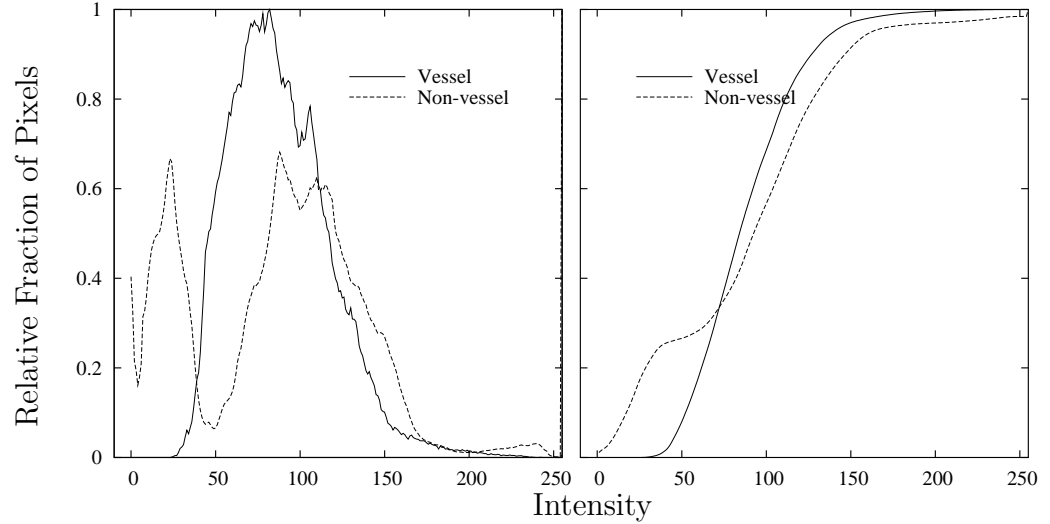
Figure 3.20: Intensity histograms of the AH dataset showing the pixel frequencies keyed by vascular status of the reference images. Cumulative intensities are given on the right.

ally' guide the development of the segmentation procedure. Note Image 5 in Figures 3.3 (page 58) and 3.34 (page 99) has extreme quality issues and bounds the minimum performance in Figure 3.21. Figure 3.22 provides the results of applying the segmentation on images not seen during development.

In order to standardize the textual representation and communication of results, the sensitivity is quoted at the matching specificity. The two trials produced classifiers that operated at the 0.85 sensitivity and specificity at a t value of 12. The initial verification dataset was small, only 5 images. This was augmented with 20 images from the (Hoover et al., 2000) study which are detailed in the next section. The standard deviations for the development data were, $s_{\text{TNF}} \approx 0.19$ and $s_{\text{TPF}} \approx 0.11$. The verification classification had greater consistentency, with smaller variations; $s_{\text{TNF}} \approx 0.13$ and $s_{\text{TPF}} \approx 0.07$ how ever the small sample size must be borne in mind.
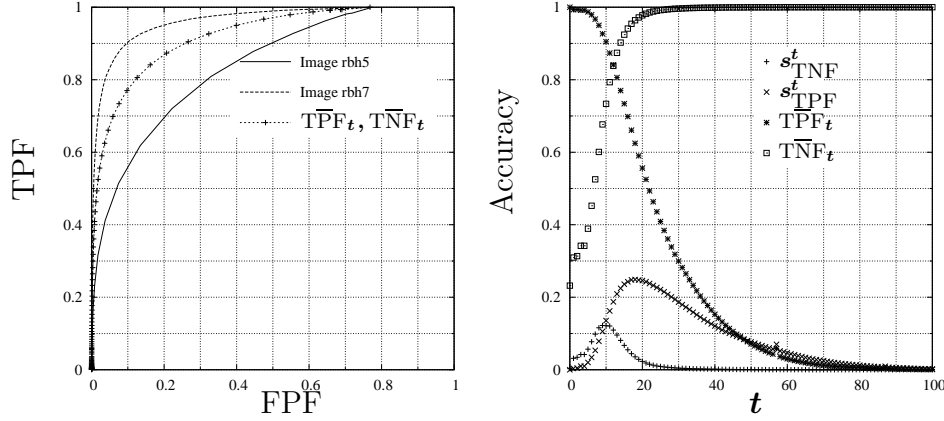
Figure 3.21: Graph on left shows Summary ROCs for RBH images 0 to 9 used during development. The ROC obtained for Image 5 bounds the minimum performance, the best operating points were obtained on image 7. The variability between the maximum and minimum range is provided by the graph on the right in terms of the standard deviation about the mean ROC. Feedback from the processing of these images was incorporated back into the development of the algorithm.

The actual numbers of pixels classified into each class is determined by their respective prevalences. For the RBH data set 5.5% of pixels belonged to the vessels, in this case 85% of these were correctly identified. On average then 14k pixels ($0.055 \times 640 \times 480 \times 0.85$) were correctly identified, 3k non-vessel pixels were identified as vessel pixels. For non-vessels picture elements, 247k of these were correctly identified where as 43k pixels were incorrectly identified as vessels. These issues are also in considered Section 3.4.2.5 of the discussion.

3.3.2   Verification results from A. Hoover dataset

The results presented in the following graphs arise from images which were not labelled by the author. In general the AH images are more detailed,
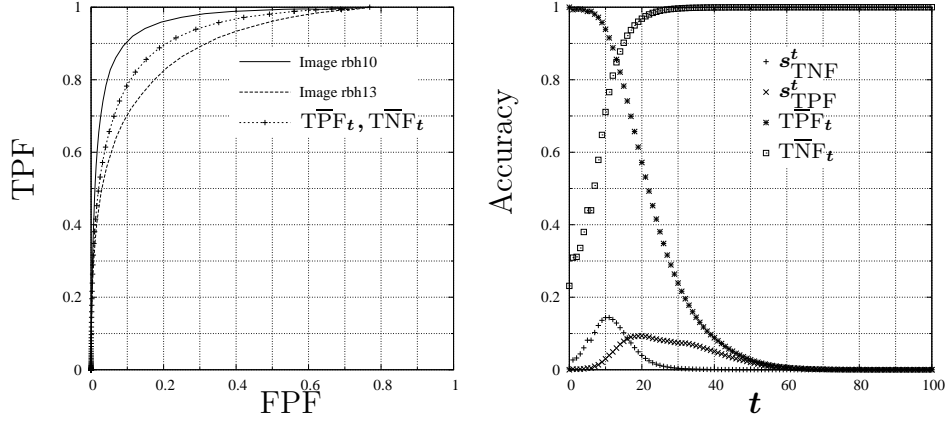
Figure 3.22: Results for RBH images 10 to 14 prospectively processed for verification of segmentation performance. The graph on the right gives the range of operating points observed, that is it bounds

because, of an increased resolution, and a smaller field of view. Also of note is the smaller proportion of images with quality issues; the AH dataset has 3 out of 20, compared to the 10 out of 15 for the RBH data set. The results of these tests can be seen in Figure 3.23. Again, by assuming equal importance between TPF and TNF a compromise operational point of $0.89=\mathrm{TNF}=\mathrm{TPF}$ can be achieved for a t value of 8. The standard deviations at this threshold values were $s_{\mathrm{TNF}} \approx 0.07$ and $s_{\mathrm{TPF}} \approx 0.14$ for all the AH images.

3.3.3 Effect of the masking operation

The masking operation of Equation 3.18 makes a large contribution to the True Negative Fraction because the majority of the reference images have very few or no vessels in the masked area. This contribution to the TNF approaches the full area of the mask, but does not equal this; as the intersection between $\boldsymbol{R}$ and $\boldsymbol{M}_{\mathrm{rbh,ah}}$ is not empty. Figure 3.24 provides the ROC and distribution of sensitivities and specificities for this case.
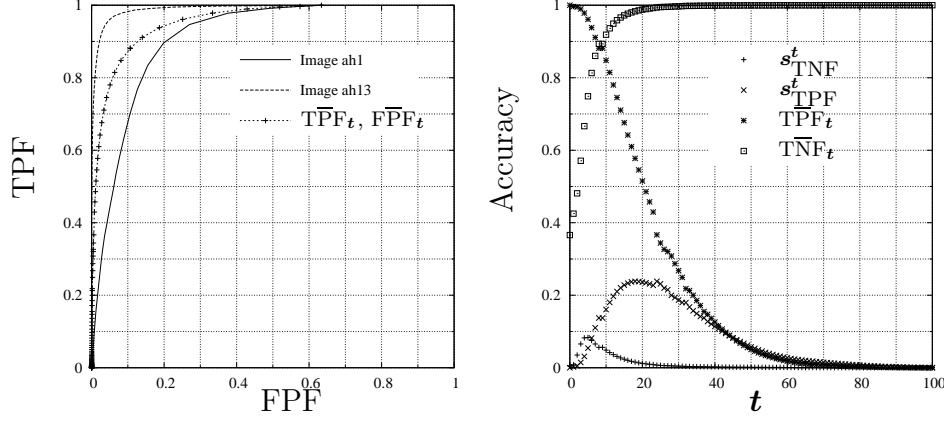
Figure 3.23: Results for AH images 0 to 19. This is the verification data, these graphs include the masking operation.

3.3.4   Results for Normal and Pathological images from the A. Hoover dataset

The results of Figure 3.25 present the AH dataset split into images with and without pathology whilst ignoring the status of the quality attribute. Only three of the images were of reduced quality with focus issues. All three of these were categorized as having pathology.

3.3.5   Quality of images

For these results the RBH and AH images have been pooled and then divided into two subsets. In one subset images with artifacts, poor focus or opacity of the ocular media, the other subset was free from these complications. In as far as was possible, pathology of the retina proper was ignored for the quality decision. Tables 3.2 and 3.1 shows this split to result in 23 images of good quality and 12 of poor quality. The ROC curves and the distribution of accuraceis are plotted in Figure 3.26.
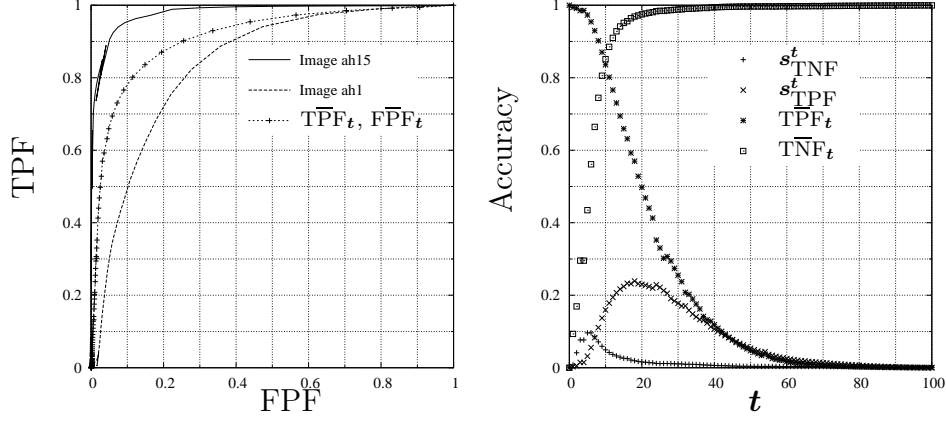
Figure 3.24: Results for AH images 0 to 19 without the masking operation. Note that in comparison to Figure 3.23 the effect of the masking operation can be appreciated by the larger dispersion of operating points of the TPF = 1 - FPF line.

3.3.6 Comparison of Royal Bournemouth Hospital and A. Hoover results

The segmentation accuracies of the whole RBH and entire AH data sets are now presented. Both the RBH and the AH images vary in resolution and field of view. The RBH subset also contains a high proportion of images with pathological or quality issues. Images 0, 2, 3 and 6 of the RBH data set being the only images without quality issues or pathology present. In contrast the AH data set which has one half of its images free from pathology or quality issues. Both these data sets were processed with the same values for r, n and B. The only difference was the respective masks used as per figure 3.11 (page 73).

The entire RBH data set was classified at a TPF=TNF=0.84 for a t value of 12. s_{TNF} and s_{TPF} were 0.10 and 0.16 respectively. The 20 images from the (Hoover et al., 2000) study were classified with TPF and TNF at approximately 0.89 for a t value of 9. The standard deviations for this data

were $s_{\mathrm{TNF}} \approx 0.07$ and $s_{\mathrm{TPF}} \approx 0.13$.

3.3.7   Comparison of images of poor quality with pathology against good quality images without pathology

Unlike the graphs of the pathology vs normal images, where the quality issue was in a 'don't care' state; the datasets in this section are split into the following two subsets: those free from pathology *and* quality issues, and those with retinal pathology compounded by quality issues. In other words, what may be the best and worst cases.

Only four images from the RBH set are free from pathology and quality considerations. Ten images from the AH dataset meet these considerations. Three images from the RBH set are considered to represent a pathological retina compounded by imaging problems (quality issues). Likewise in the AH dataset, where a further three images have both these complications. The ROCs for these subsets are graphed in Figure 3.28 on page 94.
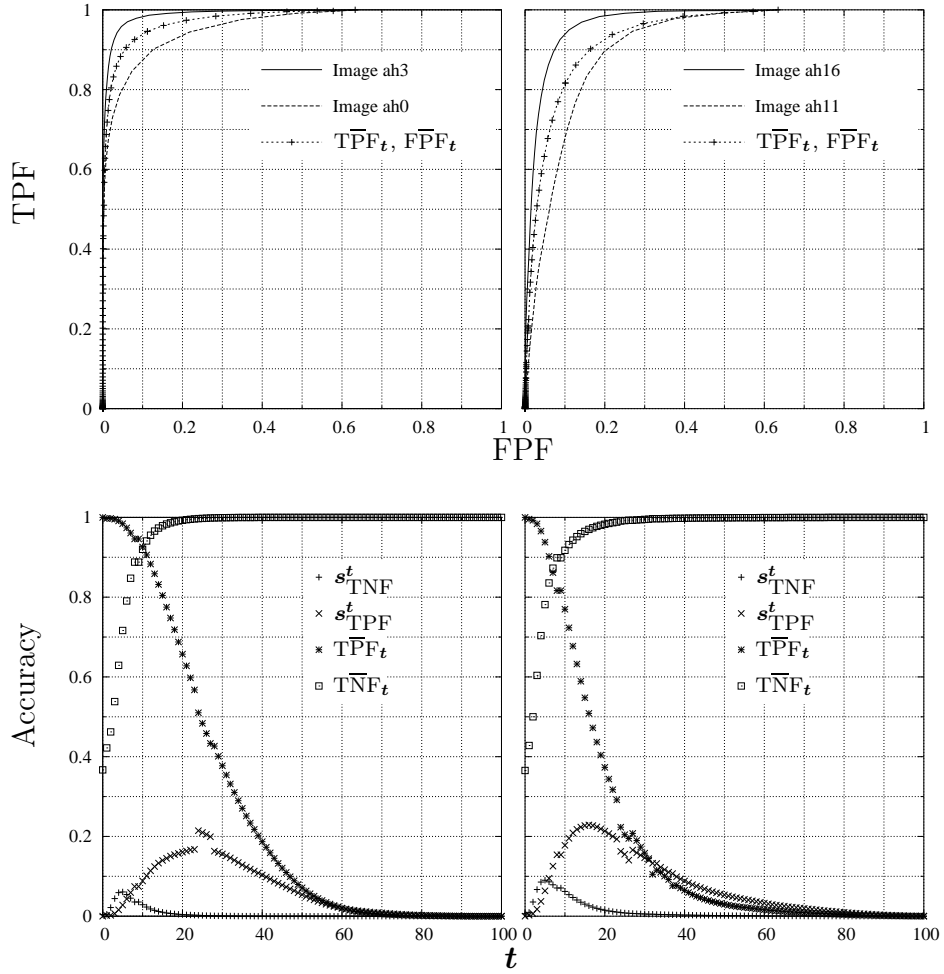
Figure 3.25: Results for AH images split into Normal (left) and Pathological (right) subsets. Any quality issues as described in Section 3.1.1.1 (page 53.) were ignored. Note the five percent (vessels and non-vessel) increase in classification accuracy of normal over pathological images
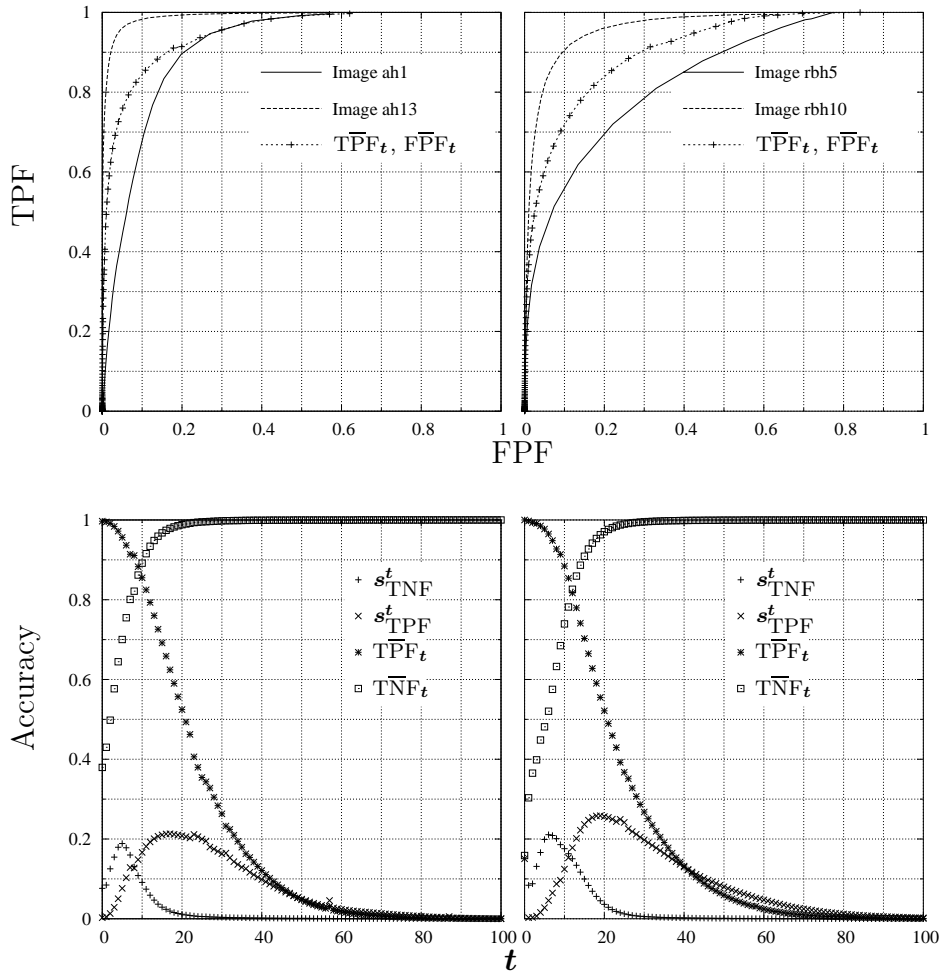
Figure 3.26: Results for all images split into 23 good quality images (left) and 12 poor quality images on right. Any pathology of the retina was ignored for the purpose of these graphs
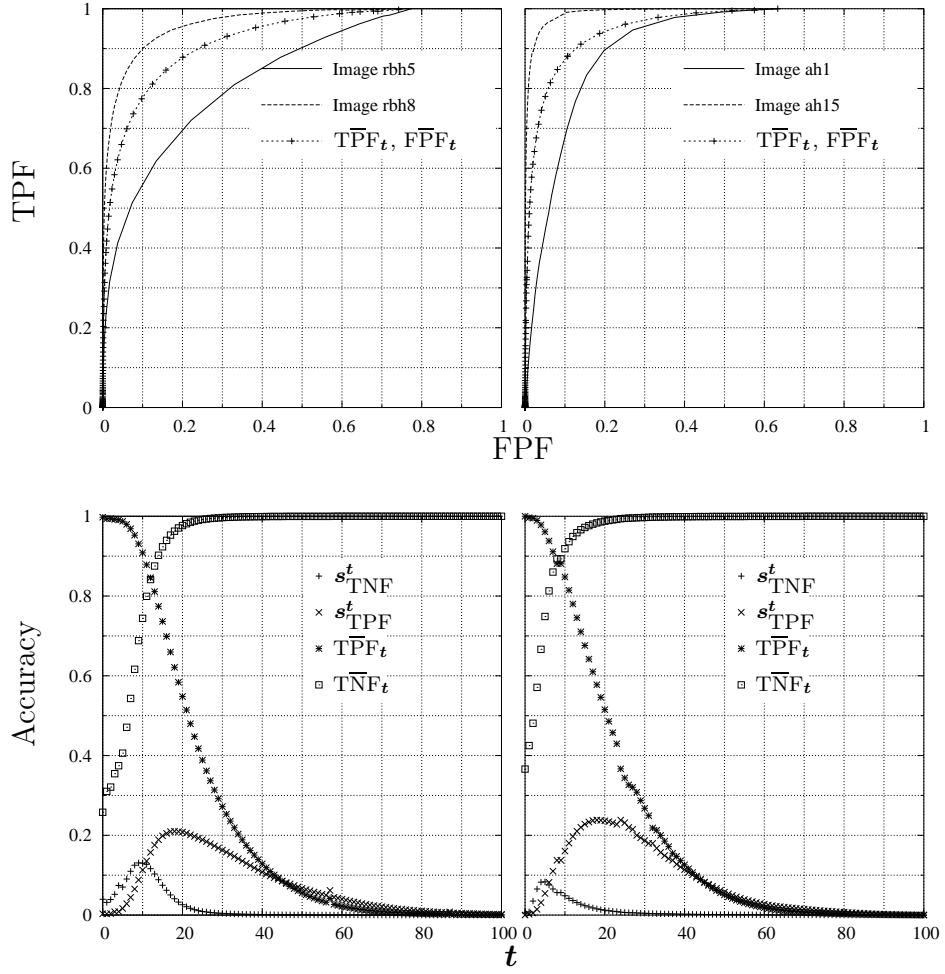
Figure 3.27: Results for all fifteen RBH images (left) and all twenty AH images (right). Note the different t values at which the mean TPF and TNF cross. For the RBH data-set this occurs at $t \approx 12$ where as for the AH data $t \approx 9$
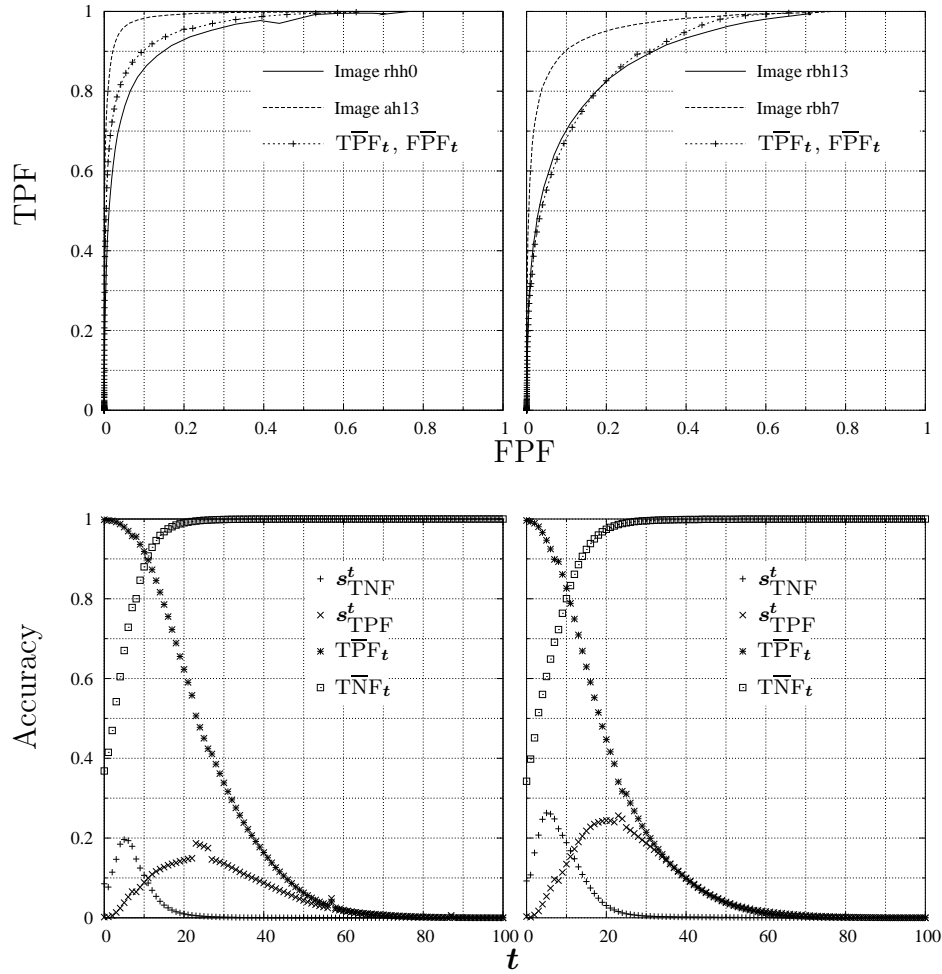
Figure 3.28: Results for the thirteen images that had no pathology and were of good-quality (left) vs pathological images which were also of poor-quality (right).

3.3.8   Sample output images

In this section instances of the output from the morphological segmentation are presented. Four images, one from each subset; normal good-quality, normal poor-quality, pathological good-quality and pathological poor-quality from either the RBH or AH data set are presented. Threshold (t) values are selected so that $\mathrm{TPF}_t \approx \mathrm{TNF}_t$ for that image. The effect of the t variable on G is presented in a series of smaller images in Figures 3.33 and 3.34, pages 98 and 99 respectively.
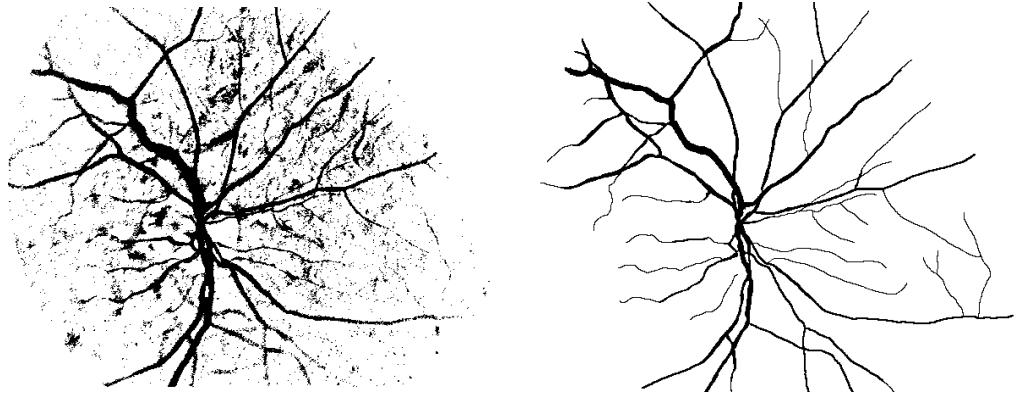
Figure 3.29: Sample output from classifier. RBH Image 8, which was considered to be of poor quality but with no retinal pathology. Threshold value t was 13. The accuracies for this image are TPF = 0.90 and TNF=0.90. The reference by which these were measured is shown on the right.
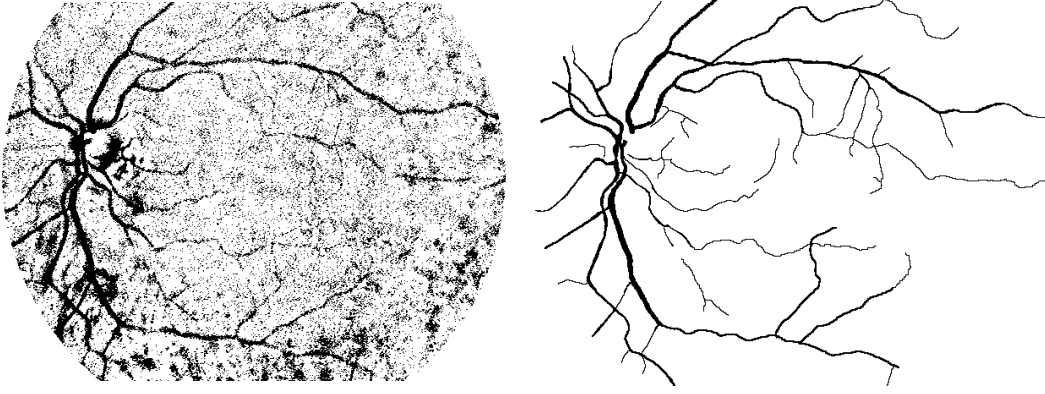
Figure 3.30: Output from classifier for an of poor quality and with pathology (Image 13). The TPF achieved for this image at $t = 18$ is 80%. The true negative fraction associated with this senstivity was 82%. The reference by which these were measured is shown on the right.
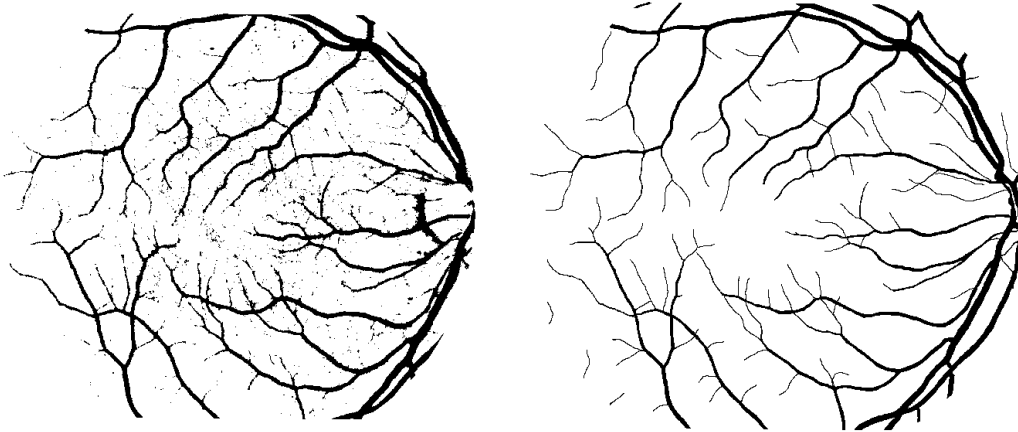
Figure 3.31: Output from classifier for Image 13 of the AH dataset. This is an image of good quality and normal appearance. The TPF and TNF achieved for this image at $t = 12$ is 95% when compared to the ground truth labelled image on the right.
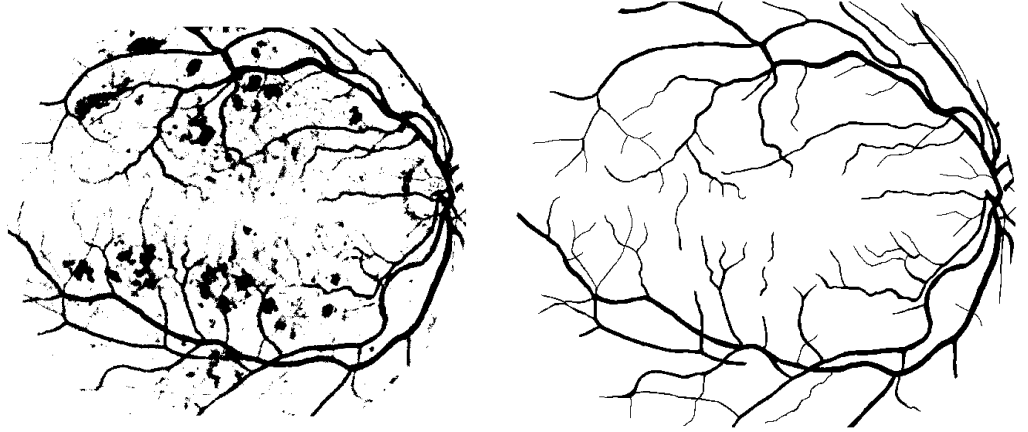
Figure 3.32: Results for an image of good imaging quality but with retinal pathology. The TPF of 93% and TNF of 91% were achieved for this image at a t value of 10. The ground truth labelled image is presented on the right.
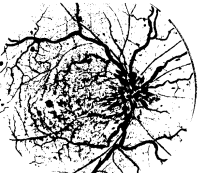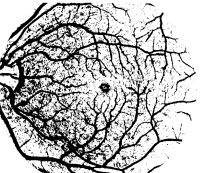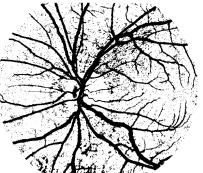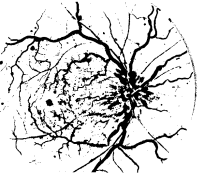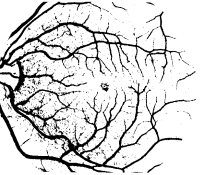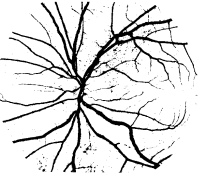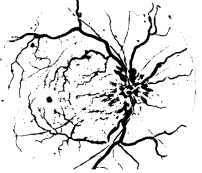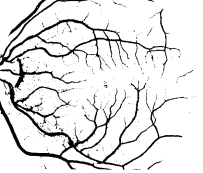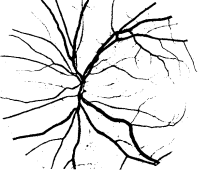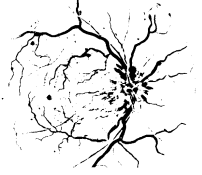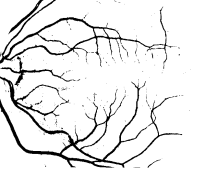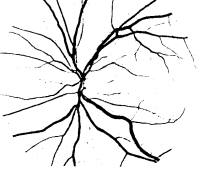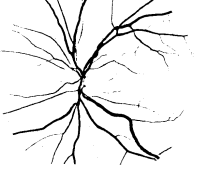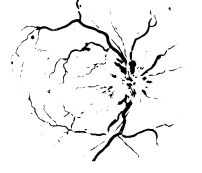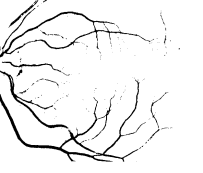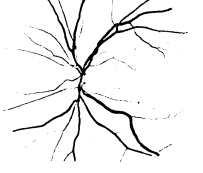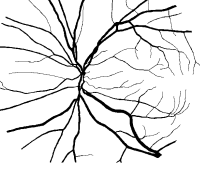
	Image			
Threshold	2	4	11	15
8				
12				
16				
20				
24				
28				
Refrence				

Figure 3.33: Sample output of the pixel classifier at various values of the threshold parameter for four images of the A. Hoover (AH) data set.
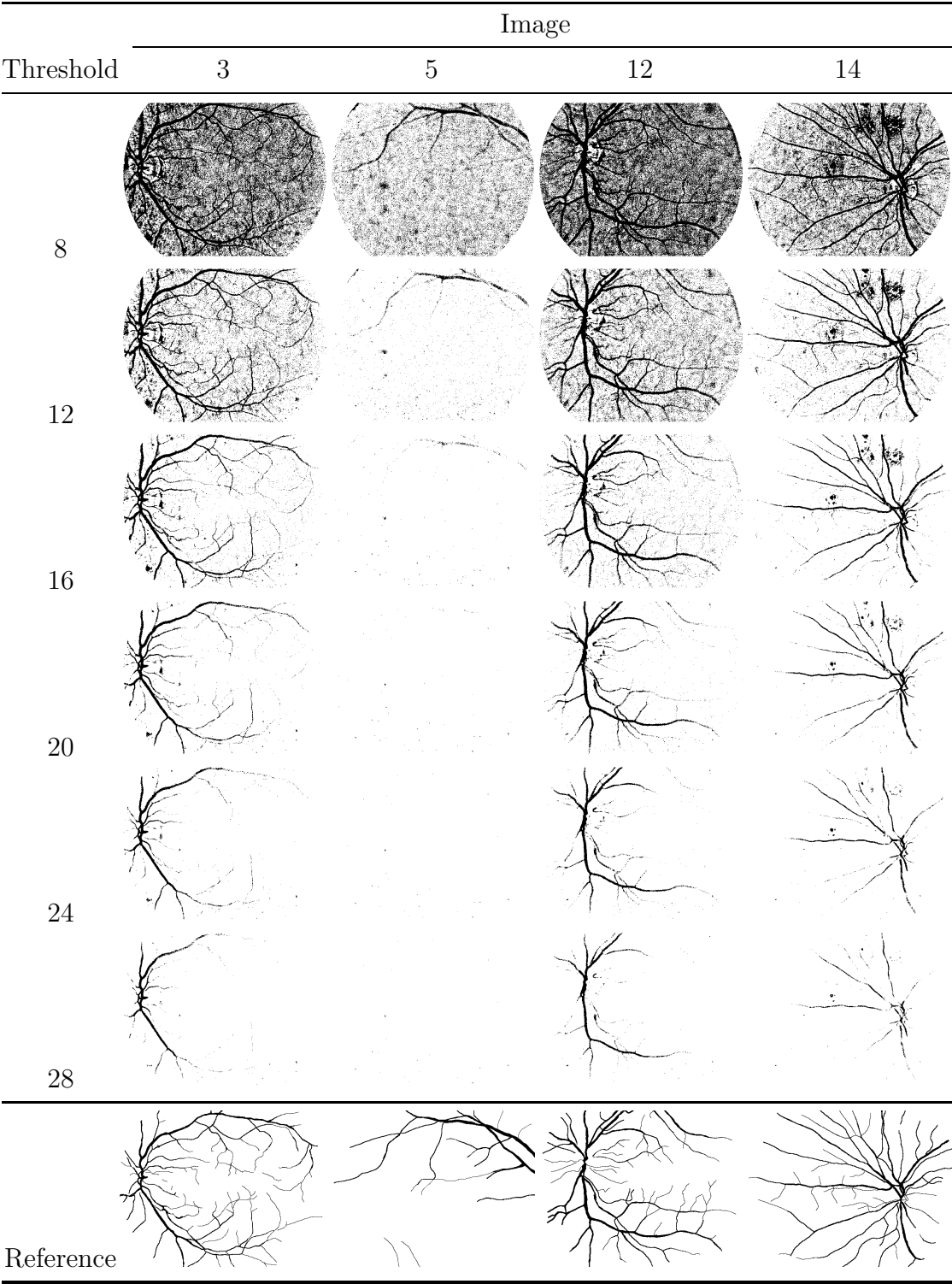
Figure 3.34: Sample results of processing various images of the RBH dataset. Note poor image 5 is of poor quality and produces little output at higher t values.

3.4   Discussion

In this penultimate section the methods and the results they produce shall be analysed. This analysis will examine the morphological algorithm, the data it operates upon and the manner in which the results are summarized. The actual results per se will be discussed and explored for observations which are of consequence or interest.

3.4.1   The morphological segmentation methods

Morphological segementation of the retinal vessels was used by Ye and Zheng (1995). However no results were given beyond a description that they were 'good'. The novelty of this research is that the segmentation method is evaluated against an idealised standard, and that the performace is well characterised in this context. This verification is the largest and most extensive test of generalisation ability known to the author. The reason for this is that the majority of research persists in publishing the segmentation results as 'before and after' processing image-pairs.

There are also some weaknesses present in the methods of the author. One of these weaknesses is in the particle analysis, which operates on the image components which remain below the threshold t of Equation 3.14. This resulted in a reduced elimination of connected-components, so that fewer than expected particles were being identified however corrective action was not considered in view of the developing results.

Another anomaly is the application of the mask M to the resulting segmented image G only. This forced the evaluation the effect of the masking, or omission of the masking, via two separate runs of the algorithm. This double run could have been avoided had the ground-truth labelled images R been masked prior to the accuracy measurements of Equations 3.19 and 3.21. The border or orientation-marker region of G could then be added to

the TNF calculation as follows:

$$\text{TNF} = \frac{\text{Card}(\boldsymbol{TN} \cup \boldsymbol{M_{\{rbh,ah\}}})}{\text{Card}(\boldsymbol{TN} \cup \boldsymbol{FP} \cup \boldsymbol{M_{\{rbh,ah\}}})} \tag{3.25}$$

This was not the case because the masking operation on $\boldsymbol{G}$ only, added an indeterminate number of FN picture elements by eliminating pixels that the algorithm had correctly identified, and thereby forcing separate runs for the accuracy measurement. In other words $\boldsymbol{M^c} \cap \boldsymbol{R} \neq \boldsymbol{\emptyset}$, or the vessels of the reference image extended into the masked region. The effects of this mismatched masking may be appreciated by observing the left image of Figure 3.31 at approximately the four o'clock position. The impact of the additional false negatives, or lack of effective particle removal is considered to be low in view of the sensitivities and specificities achieved.

3.4.1.1   Labeling

The labeling of the images was labour intensive but eased by the photo manipulation software. Unfortunately the AH images were not available to the author during development so a standardised labeling procedure was not used. Nevertheless similar datasets were produced by the author's and the Hoover et al. (2000) studies. Both studies used a binary representation for the vessel and non-vessel categories in the reference images. The case-mix of reference images was biased in the author's dataset as it was in AH's dataset; both datasets containing larger proportions of images with pathology and quality issues than can be expected in a clinical situation. On average the RBH images were labelled with 5.5% of their pixels classified as vessel. The AH images, with a smaller field of view, had 7.6% of the pixels labelled as vessel. The AH images were repeat labelled in the Hoover et al. (2000) study, the marking fidelity was found to be 90% sensitive and 94% specific.

3.4.1.2 Result summary methods

The large number of images, 35, prohibits displaying the result of each segmentation over the range of t values. Graphical summaries in the form of ROC plots were chosen because the derivation or development of a statistical or analytical model was beyond the scope of the research. The consequences of this are that inferences are taken only when data is perceived to be indicative. These conclusions, although subjective, are obvious in many cases. When the data is less clear-cut the conclusions are left to be interpreted as observations of note.

3.4.2 Discussion of results

In the following subsection a closer look is taken at the raw data as well as the segmentation accuracies derived from the labelled data. The differences between the unlabelled AH and RBH datasets are pointed out. This is followed by a discussion on the ROC curves and accuracy spreads for the development and verification image subsets. The impact on performance between normal and pathological images will be explored as well as the resilience of the algorithm to poor quality images when compared to images without acquisition problems. The comparisons cross both datasets as detailed by Tables 3.1 and 3.2 on pages 54 and 55.

3.4.2.1 Differences in the unprocessed RBH and AH image datasets

The distribution of intensities of pixels between the RBH and AH data is different. However both datasets show a significant proportion of their pixels at the zero intensity level, approximately 11% for the RBH data and 3% for the AH data. The histograms for the RBH images show an erratic distribution of pixel intensities, this is especially noticeable in Figures 3.16 and 3.19. The erratic spread of intensities over the entire range does not

seem to be affected by the split into development and verification subsets. Unlike the AH data which has a pronounced bimodal distribution (Figure 3.16 page 81) that is consistent across normal and pathological images (Figure 3.18), but not across vascular or non-vascular pixel types (Figure 3.20, page 85).

3.4.2.2   Observations applicable to all ROC curves.

All the ROCs graphs have some features in common. The most obvious of these is that the algorithm operates above the chance level for all image sets; i.e. all points lie above the TPF=FPF line. It is also clear that for masked operations on the data, the ROC curves don't span from 0,0 to 1,1 in TPF:FPF space. This is because the masking operation guarantees a significant proportion of true negatives, thus limiting the maximum FPF as already mentioned in section 3.4.1.

The use of a single image as the unit by which results are calculated has provided useful and important variability data. The use of patches, or subwindows enables (or entices) aggregation of the pixel counts, it also provides a succinct method for calculating the TPF and TNF via $\sum \text{TP}_i / (\sum \text{TP}_i + \sum \text{FN}_i)$. This type of sensitivity and specificity calculation would simplify presentation and comparisons of the data by providing a single pair of (mean) accuracies. For example, describing the specificity over 9.3 million ($\sum \text{TN}_i$) background segmentations of the entire (RBH and AH) verification subset would hide the variability of the segmentation. In effect this thesis will use this simplified method of comparing means, but it's image to image variability is also provided and fully considered.

From all the ROC curves, keyed by the threshold value t, it may be observed that the highest True Positive Fraction is achieved for $t = 0$. Likewise the false positive fraction reaches zero as t approaches 100, in effect classifying all picture elements background. In between these two extremes

of t, the response pairs $(\text{TPF}_{i,t}, \text{FPF}_{i,t})$ for individual images, and aggregated response pairs $(\overline{\text{TPF}}_t, \overline{\text{FPF}}_t)$ create ROC curves are of the expected shape; i.e. globally monotonic and concave downward (Egan, 1975).

The morphological segmentation produces variable accuracies, both for detecting vascular and non-vascular pixels. The variability of the sensitivity and specificity is dependent on the threshold value t and the image subset being segmented. The largest variability occurred for the dataset of images with poor quality combined with pathology issues, the smallest variability for the AH images without pathology, which are all free from quality issues.

The presentation of results via scattergrams or mean ROC curves are not sufficient to convey the behaviour of the morphological classifier. The mean ROCs in Figures 3.28 and 3.26 do not conform to notions of central moments or regression from the scattergram data on which it was constructed. This is because of strong and variable dependency of the accuracies of each image on t.

3.4.2.3   Effect of dataset selection

Some observations are now presented on the characteristics of the segmentation for the various datasets. Not all the possible subsets were created from the data. For example, both verification datasets were not pooled like the good-quality normals and the poor-quality with pathology datasets.

Development vs Verification The graphs in Figures 3.21 and 3.22 show little difference in the classification accuracy between the development and RBH verification images. Both datasets had a TPF $=$ TNF ≈ 0.84 at the same threshold value ($t = 12$).

For the AH verification data set the algorithm achieved TPF=FPF at a a different t value (8). Figure 3.27 on page 93 does not show the algorithm to be over-fitted the the development data as $\overline{\text{TPF}}_t = \overline{\text{TN}}\text{F}_t \approx$

0.89 for the AH data, is marginally better than $\mathrm{T\overline{P}F}_t = \mathrm{T\overline{N}F}_t \approx$ **0.84** for the RBH data. Figure 3.27 also shows a narrower range of operating points for the AH data with minimum and maximum ROC's being closer to the mean. However the difference in the variability of the operating points is mitigated by the 'nosier' composition of the RBH dataset. This lower quality and the differing fields of view between the RBH and AH datasets is also likely to explain the different t values at which the specificities equal the sensitivities.

Normal and Pathological These differences were examined for the larger AH data set. The normal images in this set had no quality issues (Table 3.1). Segmentation accuracies for the normal AH images showed the lowest variability, $s_{\mathrm{TNF}}^t = \mathbf{0.08}, s_{\mathrm{TPF}}^t = \mathbf{0.04}$, of all the image subsets. The True Positive and True Negative rates for the classification of this subset at $t = \mathbf{10}$ are $\mathrm{T\overline{P}F}_t = \mathrm{T\overline{N}F}_t \approx \mathbf{0.93}$ For reference the 2nd human labelling performed in the Hoover et al. (2000) study achieved a TPF=0.90 and TNF=0.94 for this subset.

One half of the images of the AH data had pathological abnormalities, three of these also had quality abnormalities. Using the lower two graphs in Figure 3.25 on page 91 it can be seen that for the same t value (10) and $\mathrm{T\overline{P}F}_t$ (0.92) values as the AH normals above, the ability to detect non-vessels ($\mathrm{T\overline{N}F}_t$) drops by 14% in the pathological images. In other words there is a large drop in vessel detecting ability between normal and pathological images.

Good and poor quality (pathology don't care) A similar comparison is made between images of good and poor quality. This time images from both the AH and RBH datasets were pooled; again a drop in performance of about 5% is visible in the graphs of Figure 3.26.

Best and worst case Again performance degrades as pathological and

quality complications are combined in the images. The ROCs of best-case versus worst-case data sets in graphs of Figure 3.28 on page 94, created from the pooled datasets, also show a reduction in both types of classification accuracies. The TPF and FPF is approximately 90 percent in best case images with worst case performance seven percent below this.

Note how inclusion of the RBH data into the good-quality normals lowers the best $\mathrm{T\overline{P}F}_t, \mathrm{T\overline{N}F}_t$ as images 10–19 of the AH data set, graphed in Figure 3.25, show $\mathrm{T\overline{P}F}_t, \mathrm{T\overline{N}F}_t$ pairs which are consistently greater than the aggregated super set.

To summarise; it can be seen that classification of individual images (within data subsets) creates significant variations in the $\mathrm{TPF}_{t,i}$ and $\mathrm{TNF}_{t,i}$, yet their mean values ($\mathrm{T\overline{P}F}_t, \mathrm{T\overline{N}F}_t$) show a smooth progression. It also suggests normalizing (via a histogram stretch or shrink for example) the intensity of images c, to increase the consistency of desired classification accuracies as controlled by t across datasets.

The images used in this study were taken using different fields of view and resolutions. Despite these favourable characteristics of the dataset, a resolution and scale dependence is likely to remain. The effects of this scale dependence are also the likely cause of the difference in threshold (t) values at which the RBH and AH classifications attain $\mathrm{T\overline{P}F}_t = \mathrm{T\overline{N}F}_t$.

As sensor technology improves both the radius r of equation 3.12, the connnected component size n, and the size of B will need adjusting. This adjustment could be as simple as making r, n and B scale linearly with the resolution.

3.4.2.4 Effects of the masking operation

The masking operation adds a significant number of true negatives to the overall classification. Ideally this would not be included in the results, but
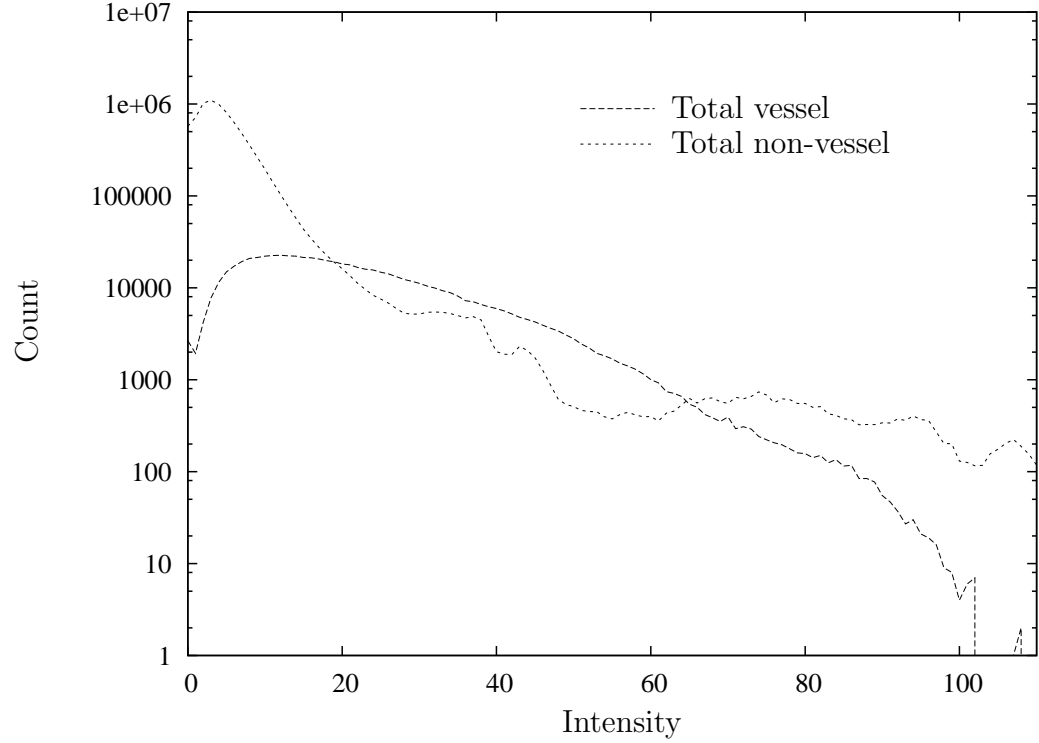
Figure 3.35: Vessel and non-vessel pixel counts for the background sub-tracted image c, taken from the AH dataset. This graph shows the degree of separation between the two pixel types.

to ensure consistency with the presentation of results from other studies, sensitivities and specificities are not given for regions of interest, but instead for whole images including the masked region. See Figure section 3.11 and the short discussio in Section 3.4.1 on page 100.

3.4.2.5   Comments on display of processed images

The classified images displayed at t values so that $\mathrm{T\overline{P}F}_t = \mathrm{T\overline{N}F}_t$ are not as aesthetically pleasing as images at higher threshold values, where there are less false positives. Compare the Figures 3.34 (page 99) and 3.33 (page 98) at $t = 16$ or $t = 20$, with the images on pages 95 and 96.

Interestingly the range of t values for which results are visually most appealing is also the range of t values where the probability of correct classification is highest. Results have been reported at thresholds so that the classification rates for non-vessel pixels equal those of vessel pixels. This reporting method communicates a general impression of the ROC going from 1,1 via TPF,TPF|TPF=TNF to 0,0 constrained by the fact that its gradient is non-decreasing. It is also customary to give the results at this point as no costs are associated with misclassification of either class.

This point on the ROC, where TPF=TNF, is not necessarily the point that gives greatest overall proportion of correct responses. That is because the prior probabilities, or the prevalence of false (non-vessel) cases is not equal to the prevalence or number of true (vessel) cases. Thus one type of classification accuracy has a disproportionate effect on the total classification. In other words, the total correct fraction CF:

$$\text{CF} = \sum_i \text{TP.T}\overline{\text{P}}\text{F} + \sum_i \text{TN.T}\overline{\text{N}}\text{F} \tag{3.26}$$

For the AH dataset the TP count = 0.076 (pixels labelled as vessel), therefore the and TN count = 0.924. The $\text{T}\overline{\text{P}}\text{F}_t$ and $\text{T}\overline{\text{N}}\text{F}_t$ are obtained via the morphological operations. In the plot of CF vs t presented in Figure 3.36 the maximum correct fractions, of 94% and 96% for the AH and RBH datasets are achieved in the region of $t = 20$. It is interesting to note that the TNF is almost 100% and the TPF is 50% for this threshold.

3.4.2.6   Further work

There are a number of issues concerning the method and the analysis which could be improved. Out of the many enhancements and developments that are always possible, the most desirable is to find an automated method to close the feedback loop. In other words coupling the classification accuracies to the parameters that control the algorithm. Genetic algorithms are
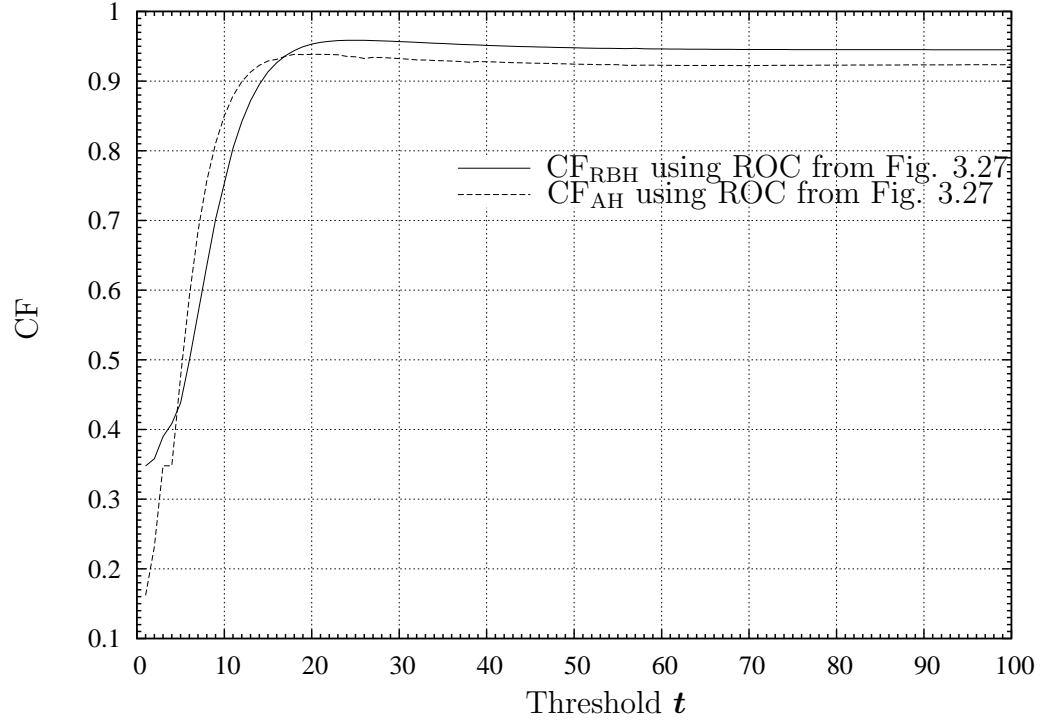
Figure 3.36: An alternative presentation of the vessel segmentation classification accuracy. It illustrates the small contribution correct classification of the vessel pixels.

well suited for this type of optimization. A brute force method should also be considered because there are only three main parameters n, r and B controlling the operation of the segmentation, although many more could be created.

Selection of a threshold value is of practical interest and is dependent on the purpose of the segmented image; for visual inspection t could be selected so that the segmented (TP + FP) matches the prior probabilities, or prevalence of vascular pixels. If spectographic analysis is to be undertaken of retinal tissues a max(CF) t value would be more appropriate.

Further improvements in the rates at which the non-vessel and vessel pixels are identified could come from investigating whether the background

subtraction of equation 3.13 a sphere is indeed the optimum shape. Would taking the maximum response of a 'tumbling ellipse' improve the profile presented in Figure 3.10 (page 70) ? Further work would also include investigating if unsharp masking (the RBH dataset in particular) is beneficial. Future work is also likely to include the elimination of the intermittent, yet detracting, glitches that can be seen around $t = 25$ in the graphs. Switching from the rapid application development environment to a more stable software platform would accomplish this.

The most immediate and practical improvements needed by this work include adjusting the mask operation, and correcting the particle analysis. To further demonstrate that this method is robust in the face of dataset selection, or to bolster the applicability and representative nature of the results, larger datasets of images could be used. A second labelling of the RBH images, and a third labelling of the AH images would also be considered.

3.5   Conclusions and Chapter Summary

This chapter concludes by highlighting the strengths, improvements and conclusions applicable to morphological segmentation. In the context of the work presented in Chapter 2 the approach presented is relatively simple. It is also tested and characterised upon a more diverse and difficult set of images than any other study. This means that the method is potentially suited for use as a reference test when more sophisticated or elaborate vessel segmentation algorithms are developed and published. The classification rates of morphological segmentation are similar to, if not above, those few publications that have reported their results (Hoover et al., 2000; Sinthanayothin et al., 1999). The main drawback of this approach is that it doesn't incorporate structural assumptions about the vessels, so

the results may be described as vessel-pixel orientated rather than focused on the secondary or more abstract geometry of vascular arbour.

The parameters selected for the morphological segmentation based on the RBH data did not tailor the performance to the development dataset. However this does not demonstrate the scalability of the morphological segmentation. Images captured at widely different resolutions and modalities would require testing to ascertain this. The question whether there are any tenable inferences to be made on the applicability of these results to a wider 'real world' can be partially answered. For reflectance fundus images, with fields of view (20°–50°) and resolutions less than 0.5 million pixels typical of the 1990's decade, morphological segmentation is successful at the pixel level.

The segmentation strategy is simple and well characterised in its performance. It possesses a high degree of mean accuracy both in terms of specificity and sensitivity. This makes it well suited as a reference when developing novel strategies, or when it is necessary to identify vascular or non-vascular picture elements with the greatest possible accuracy.

This page intentionally contains only this sentence.

Chapter 4

Artificial neural network segmentation

4.1 Introduction

Artificial Neural Networks (ANN) have been demonstrated as effective classifiers in medical pattern recognition (Baxt, 1991a; Baxt, 1991b; Page et al., 1996) and have become viable methods for image processing (Egmont-Petersen et al., 2002). In this thesis, the neural network performs a segmentation directly on pixel-level data without preprocessing. As in the previous chapter the aim of developing the neural networks is to produce classifiers for the identification of vascular picture elements. The following chapter will detail the reasons, methods, results and evaluation of a large ANN based on pixel level data.

ANN classifiers have advantages in ease of use and in the ability to learn from examples, in contrast to much of 'traditional' image-processing, in which the results of processing are presented as qualitative and subjective images. Artificial neural networks trained using supervised techniques have the inherent benefit that the availability of objective target

data provides. This data enables quantitive assessment of performance during training, but more importantly it also facilitates prospective assessment.

The choice of methods employed by the authors have much in common with other projects using ANNs, in that they are relatively ad-hoc and guided by rules of thumb. Feed forward networks using the back-propagation of errors for weight corrections were chosen. The reasons the back-propagation networks were chosen is the small number of parameters and their 'established' nature. In addition to these advantages this academic area is novel, only two previous studies have identified the fundus vessels using ANN technology, see review in Section 2.6.4.

One these two studies (Sinthanayothin et al., 1999) was designed to identify individual pixels the other (Gardner et al., 1996) was programmed to detect image sub-windows containing vessels. Both detectors had similar performance. The pixel detector had a true positive fraction of 83%, the true negative fraction was 91%. Sub-windows twenty pixels square were classified as containing vessels or not with a specificity and sensitivity of 91.7%.

Sinthanayothin et al. (1999) used 112 images drawn from an imaging clinic. Pixel-level data was used to form inputs to the neural network. Two hundred inputs in total, one hundred of which were derived from an edge detection applied to a ten by ten pixel subwindow. Another one hundred inputs were taken from the first band of a principal component analysis of the same ten by ten subwindow. The composition of the dataset from which the images were taken was not specified, however if it was typical of an imaging clinic approximately 10% of images can be expected to be of poor quality (Kristinsson et al., 1997; Klein et al., 1989b; Liesenfeld et al., 2000), and a similar proportion can be expected to have mild to severe NPDR diabetic pathology (Klein et al., 1984). A similar proportion again can be expected to have photocoagulative scars (Klein et al., 1989b). Al-

together this is a significant level of potential interference and does not include co-morbidity, or pathology not associated with DR.

In general as computing resources are limited, much of the neural-network design process is usually dedicated to pre-processing the data so as to reduce its dimensionality. If this reduction of dimensionality is lossy, i.e. information presented to the network cannot be recreated to its original form, the designers run the risk of ignoring potentially relevant data. This task of re-representing the data is called feature selection or feature extraction, and it is potentially so problematic it is an area of research in it self (Prakash and Narashimah Murty, 1995; McNitt-Gray et al., 1995; Sahiner et al., 1996)

With, exception of the mathematical internals, ANNs lack a formalism or engineering guidelines. Despite this, there is a wealth of experience and heuristic knowledge in the form of previous pattern recognition research. Many of the choices for the structure, algorithms and the parameters for the network were a combination of examples drawn from research and a process of trial and error.

The ANN designed for this research is directly connected to minimally pre-processed fundus image. This arrangement bypasses the need to develop features and assess them for relevance. However many of the usual requirements of a neural network datasets are still present. The data has to encapsulate or fully present the underlying mechanisms that generate it. With a directly connected neural network this requirement is somewhat alleviated, however during production of the 'gold standard' training and verification data, more information was used by the grader than is available to the network. The grader had the benefit of the context of the whole fundus, as we shall see in the next section, the network has but a subwindow.

A similar requirement stipulates that enough data be present to 'delineate' the boundaries of the classes in pattern space. In other words, given a fixed number of input patterns (from the which features are drawn) as

one increases the input dimensionality (or number of features) to describe each pattern, the poorer or sparser the distribution of the patterns in patterns space is defined. This is often referred to as the dimensionality curse. In this study the author chose to overcome this requirement with a brute force approach, with four orders of magnitude more data training samples than input dimensions.

The use of features can present the data to the ANN in a scale, rotation and translation invariant fashion. A directly connected ANN may not generalise because these transforms are not explicitly incorporated into training. These transforms can either be encoded within the network or artificially introduced into the training and verification data in to ensure that ANN classification behaviour is not dependent on the scale, rotation or translation of the input. These are usually concerns when there is a risk that the training data is not representative of the population data. In this study the author makes the assumption that enough samples of rotated and scaled vessels and translated non-vessels were represented by the dataset itself, to ensure adequate generalisation.

4.2   Methods

This section will detail the data used, the topology of the network and the training strategy used.

4.2.1   Data preparation and sampling

The data used in the authors study is described in Chapter 3. The datasets used in the training and verification of ANN were prepared from both the RBH and the AH datasets. The division of the dataset into training and verification (generalisation) datasets is shown in table 4.1 for the RBH data, and in Table 4.2 for the AH data.

Image Number	Pathology Present	Poor Quality	ANN Verification
0			●
1		●	
2			
3			
4	●	●	●
5		●	
6	●		
7	●	●	
8		●	●
9		●	
10		●	
11	●		
12		●	
13	●	●	
14	●		●

Table 4.1: Training images in Royal Bournemouth Hospital (RBH) dataset, images not used were reserved for verification.

From the green plane of these images the training and verification samples were drawn as follows. The training data was created by randomly sampling the vessel and non-vessel pixels from the images in Tables 4.2 and 4.1. However as vessel pixels only represent $\approx 6\%$ of an image, in the training data they were oversampled to represent 50% of training samples, however only 1.3% of the 600k vessel samples were sampled twice.

The intensities of the original images, $\boldsymbol{a}$, were rescaled onto the $[\boldsymbol{-1, 1}]$ range, so that they preserved their original relative brightnesses:

$$s = \frac{2a}{N - 1} - 1 \qquad (4.1)$$

where $\boldsymbol{N} = \boldsymbol{255}$ and $\boldsymbol{s}$ is the rescaled image. The target outputs during training were set to 0 for non-vessel picture element and 1 for a vessel pixel. In this manner, 1.2 million training samples (0.6M vessel samples and 0.6M non-vessel samples) were created.

Label	Number	Pathology	Quality	ANN Verification
im0001	0	●		
im0002	1	●		
im0003	2	●		
im0004	3	●		●
im0005	4	●	●	
im0044	5	●		
im0139	6	●		
im0319	7	●		●
im0240	8	●	●	
im0324	9	●	●	
im0291	10			●
im0077	11			
im0081	12			
im0082	13			
im0162	14			●
im0163	15			
im0235	16			
im0236	17			
im0239	18			●
im0255	19			

Table 4.2: Verification images in AH dataset, images not used were reserved for training.

The degree to which any of the pixels had areas in common is now estimated. The samples came from the 26 images used in training. As detailed in the next section the ANN had an input dimensionality of 41^2, meaning the available sampling area to the image is reduced by 40 pixels in each dimension. Thus from the 11 RBH and 15 AH images an area of 2915 pixels squared is left from which to sample. Given the ≈ 0.067 prior probability that a pixel is a vessel, 2816 pixels squared is thus the area which can be sampled to create non-vessel training data. From this area 0.6M samples of must be drawn. In other words, $\mathbf{0.6M/2816^2}$ or each pixel has a 0.075 chance of being included as a target, In terms of area; one in every 13.33

pixels is used, i.e. one sample is on average 3.65 pixels distant from the next in either x,y or x and y directions. In reality it is likely that there is less overlap, or greater distance between samples as the non-vessel data is interspersed with vessel data.

The sampling density for vascular pixels is different. There are 592k vascular samples to choose from but 600k are used. The result is that vessel input patterns overlap with their neighbours to a great extent and that small number (8k) have identical targets, i.e. they are included twice. Given that every vessel pixel is used, each pattern will be substantially represented by one or more of its neighbours.

Separate datasets were used to monitor the progress of training. These were created by sampling the verification images. Five images from the AH dataset and four images from the RBH dataset provided 189k vessel pixels ($\mathbf{5 \times (575 \times 660) \times 0.076 + 4 \times (600x440) \times 0.055}$) and 2.7M non-vessel samples. From these totals smaller subsets containing 7k vessel patterns and 7k non-vessel patterns provided verification MSE in training. However in final testing, and construction of the ROCs, every pixel (bar twenty pixels near each border) of the verification data was used.

4.3   Artificial Neural Network Architecture

During labeling of the reference or 'gold-standard' images it was observed that the process was facilitated by the context. Seeing a portion of the fundus aided identifying the vessels. This is easily appreciated if one attempts to categorise a subwindow that is composed of a single pixel. The single pixel provides little of relevance other than prior probability (if available), a two by two sub-window can start to give some gradient information, yet it is still subject to noise. Increase the window size further and it progres-

sively becomes easier to label the centre pixel as belonging to one of the classes.

The desirability of having a large input window was offset against the dimensionality curse and computational resources. From this compromise emerged a selection of an input dimensionality of 41 picture elements squared, i.e. 1681. The centre pixel of this square subwindow is the one classified as vessel or not.

The target pixel is also the single output from the network. The network, in prediction mode, can be visualised to operate as a raster or 'sliding window' over the image. No classification was performed within a 20 pixel border of the edges, the missing values were not imputed. Picture elements in these borders have been excluded from all subsequent analysis. The majority of these pixels belong to the orientation mask and it is assumed they have little or no impact upon classification rates.

The number of internal or hidden layers of the network was chosen to follow the proportions used in previous work (Eberhart et al., 1991; Wu et al., 1993; Baumgart-Schmitt et al., 1997; Bhattacharya et al., 1997). These suggested hidden nodes numbers to be a fraction of the input dimensionality. Forty nine hidden nodes were chosen in keeping with the general proportions observed. The hidden nodes were fully connected to the inputs and in turn formed all the inputs to the single output node. The total number of weights was thus 41x41x7x7+7x7, or 82418.

4.3.1 Artificial Neural Network: Algorithm

The macro language of the Stuttgart Neural Network Simulator (SNNS) [1] software was used to implement the ANN. The delta rule for error backpropagation was used for weight adjustment (Bishop, 1995). Before training, the weights were randomised between values of -1 and 1. Then during

---

[1]http://www-ra.informatik.uni-tuebingen.de/SNNS/

training weight corrections were accumulated or cached over N_w patterns before being updated. This weight update was scaled by factor of η (learning rate). No weight update was performed if the difference between target and network output was smaller than d_{max}. Trial runs were used to determine the settings for these parameters.

4.4   Training

In training the patterns were presented to the network and mean squared error (MSE) and verification MSE (vMSE) were observed. The two main problems during training were meeting processing and memory requirements. Processing demands were met by verifying convergence within the first few training epochs using a range of values for η and d_{max}. The memory requirements were met by staggering the loading of samples into SNNS memory.

The staggering of training samples means that an epoch refers to a batch and not the whole sample set. The spikes that are visible in many of the vMSE and MSE traces are also the result of processing differing batches.

4.4.1   Trial runs

The chunk size variable (N_w) was set at 50 and a range of d_{max} and η were subjected to a brief examination. Using 20k samples the d_{max} parameter was varied over the 0.0–0.075 range for learning rates of 0.05, 0.10 and 0.15. A similar trial was run with 34k patterns this time using a range of 0.025–0.100 for the d_{max} parameter and the learning rate set to 0.10, 0.15 and 0.2. The results are displayed in Figures 4.1 and 4.2.

Using an AMD Duron processor, running at 800MHz, the neural networks were observed to process approximately 5700 samples per minute.

(a) $d_{max} = 0.00$

(b) $d_{max} = 0.025$

(c) $d_{max} = 0.050$

(d) $d_{max} = 0.075$
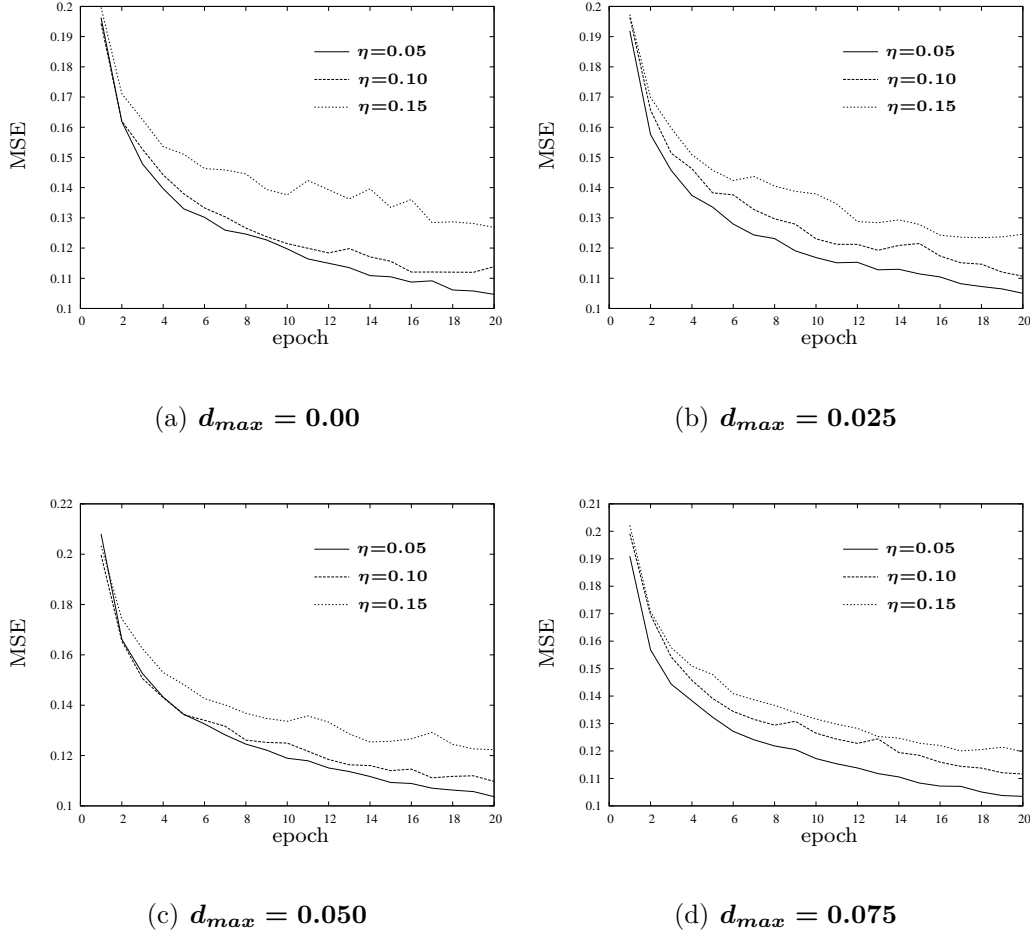
Figure 4.1: Trial training runs investigating the effect of η and d_{max} on convergence with a training dataset of 20k samples.

However this figure does not take into account considerable overheads such as loading the batches or processing the verification data. The 600k patterns used in the final training were divided into 10k-sample batches. A batch loaded into memory was considered to represent an epoch; so that 60 epochs are necessary to present the full training set to the network.

In the trial runs, with the smaller sample sets, it was observed that lower values of η, and higher values of d_{max} resulted in more reliable convergence, training using the full data set was started.
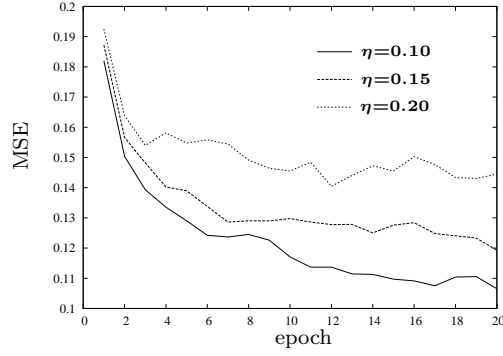
(a) $d_{max} = 0.025$
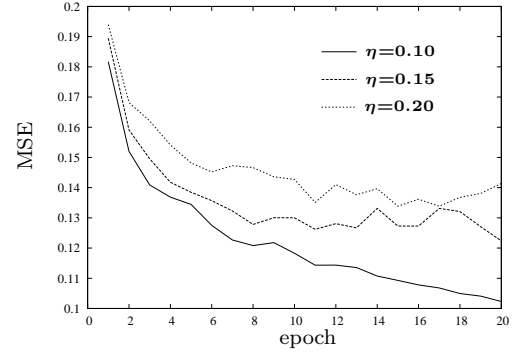
(b) $d_{max} = 0.050$
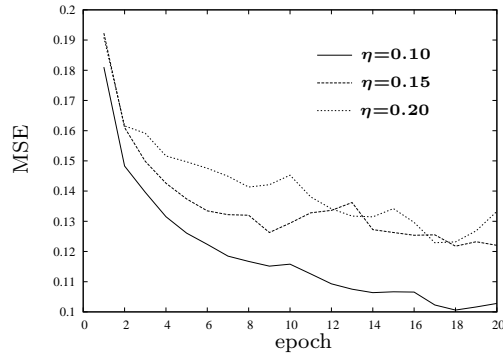
(c) $d_{max} = 0.075$
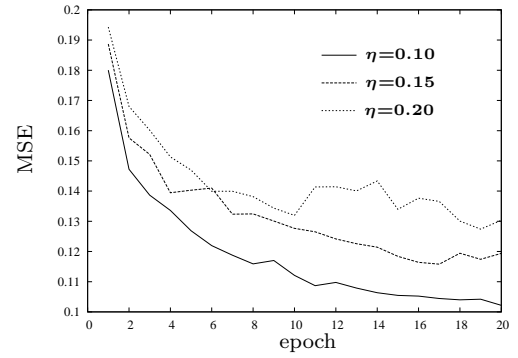
(d) $d_{max} = 0.100$

Figure 4.2: Training runs evaluating the effect of η and d_{max} on convergence over a 34k pattern set.

Figure 4.4 shows the error rates for the vMSE and MSE for a longer training period. Note at 180 epochs the vMSE and MSE errors appear to start to diverge, and that further training could be counter productive. To estimate the point at which further training was, no longer, or counter productive, training was extended to 440 epochs (See Figure 4.4). A similar training run was performed over 360 epochs using a learning rate of $\eta = 0.05$ and d_{max} values of 0.025, and 0.050, this is shown in figure 4.5.

The longer training runs showed that there was no benefit to training
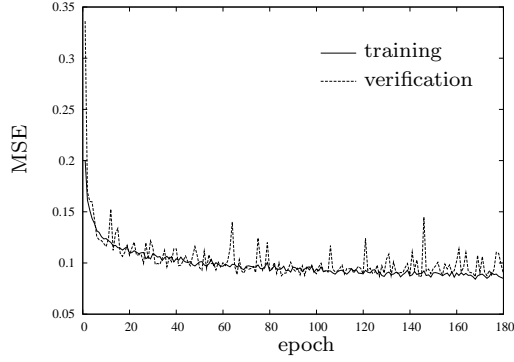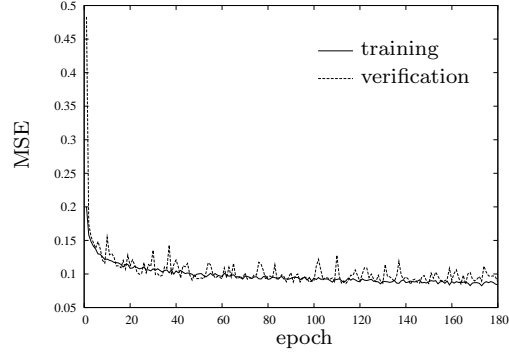
(a) $d_{max} = 0.025$ (b) $d_{max} = 0.050$

Figure 4.3: Network errors over longer training and over larger datasets (600k samples) and η=0.05, each epoch represents a 10k batch of samples.

beyond 180 epochs. A final two networks were trained for 180 epochs using learning rates of 0.005 and 0.01, their error rates are illustrated in Figure 4.6

From these two and previous trainings instances of the networks with low vMSE were assessed as classifiers.
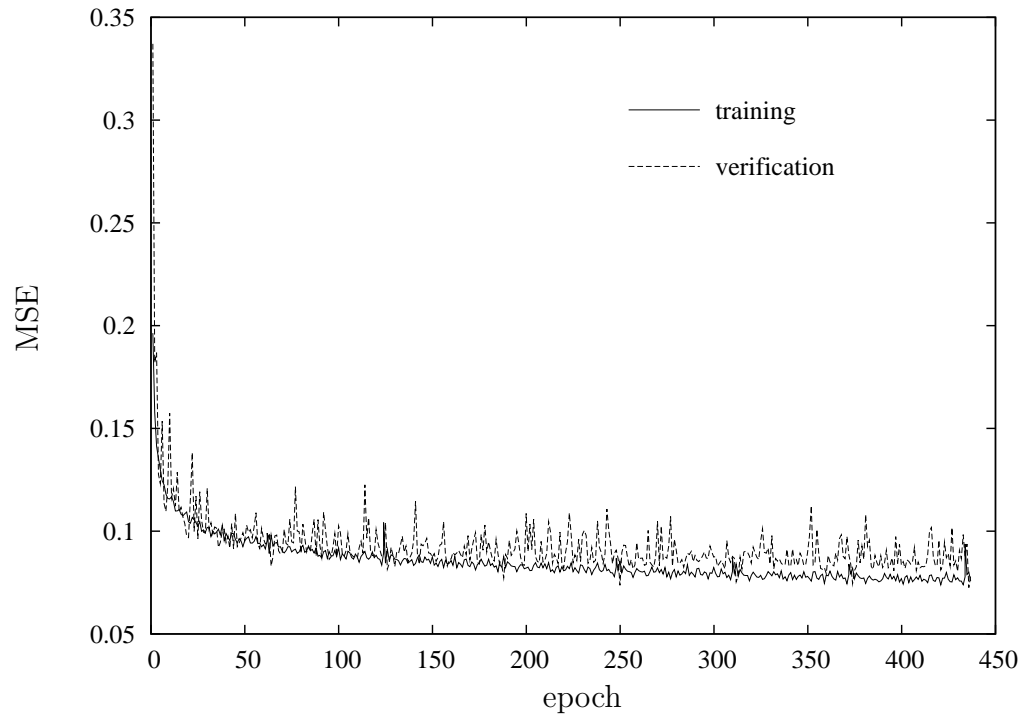
Figure 4.4: Training for 440 epochs, each of a 10k batch, using $d_{max} \quad =$ $0.025, \eta = 0.035$
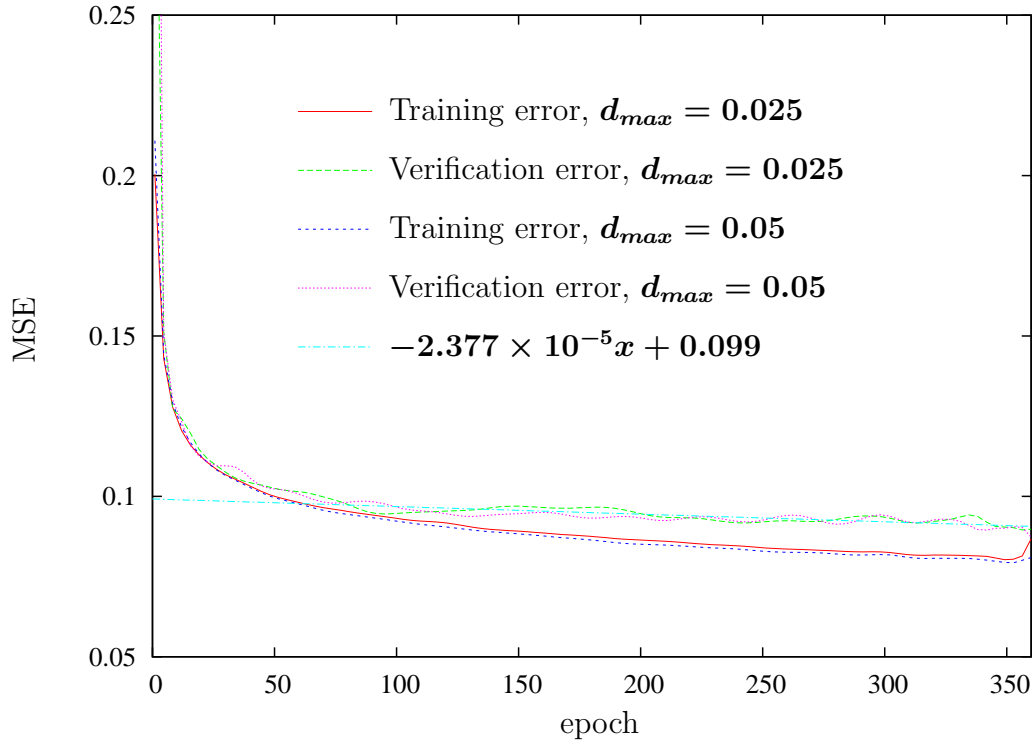
Figure 4.5: Smoothed errors, from training the neural network on the full dataset using $d_{max} = \{0.025, 0.050\}$ and $\eta = 0.05$. A regression line has been fitted to the verification MSE between epochs 150 to 360.

(a) $\eta = 0.005$

(b) $\eta = 0.010$
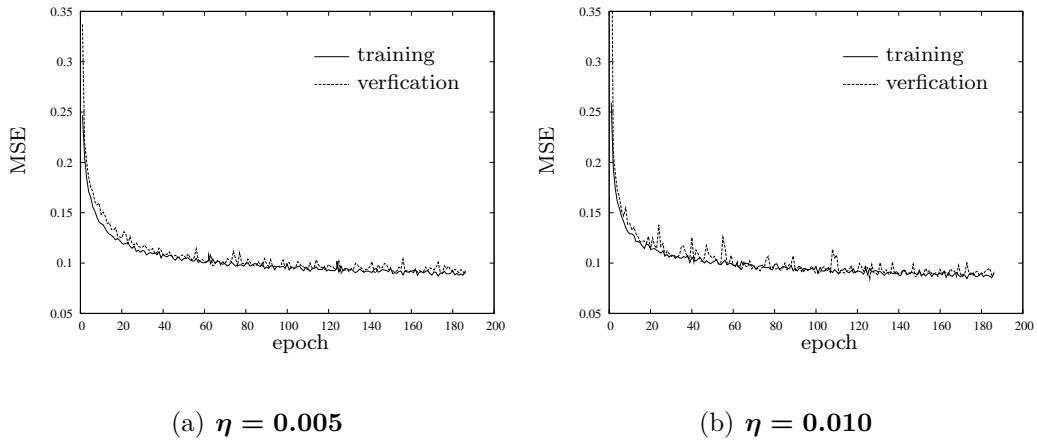
Figure 4.6: Final network training attempting to lower the vMSE

4.5   Results

The verification of performance by processing the images (detailed in Tables 4.2 and 4.1) is presented in this section. The neural networks when presented with these images as input, output images in which each pixel represents the response or measure of evidence for the presence of vessel. These response images are thresholded to obtain a binary classification of the vasculature.

The results of neural segmentation of the vasculature are presented in the same manner as the morphological segmentation results; the output of the neural networks was compared to the 'ground-truth' labeled images. ROC's are plotted using the range of performance tradeoffs obtained over a range of thresholds. These ROCs are given for individual images as well as aggregated pixels. Furthermore ROCs are displayed for networks of various vMSEs that span the range of vMSEs obtained during training, as well as the results of processing the images they were trained on.

The ROCs better characterise the distribution of the errors than the vMSE, as the thresholding distinguishes whether the error occurred for a vessel or non-vessel classification. To maintain consistency of presentation the threshold range, although spanning the full intensity range of the output images, was remapped onto the [0:100] interval.

The first result to be presented is the ROC for the network with the lowest vMSEs achieved. This was 0.0705; the trade off in vessel against non-vessel segmentation accuracy is shown in Figure 4.7. In theory this network has the lowest prospective error rate and thus the best classification ability. However the noisy or irregular nature of the errors during training requires that networks of other low vMSEs are presented.
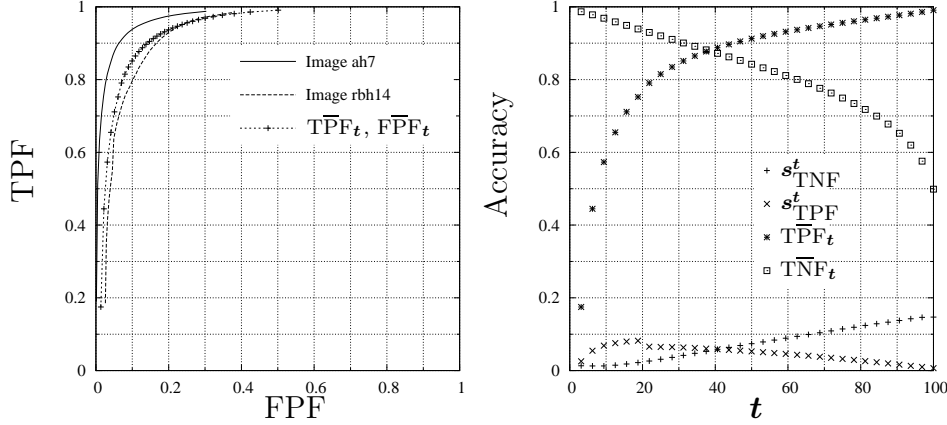
Figure 4.7: Operating Characteristics for a network of vMSE=0.0705, tested on 9 verification images and compared to human ground truth labeling. The graph on the left shows the best (top solid line), mean (dotted with cross points) and worst (dashed) labeling behaviour of this network. All other results are bounded by the best and worst lines.

To establish the relationship between vMSE and the corresponding ROC, networks at various vMSEs are investigated. These networks were obtained from alternate training runs using different initial weights, or from earlier in the training cycle. Out of curiosity the ROCs for the training datasets, which reported consistently lower MSE after 180 epochs are also plotted. Table 4.3 summarises the remaining results presented in this section.

Figure	vMSE	Description
4.8	0.0815 and 0.0901	Networks with low vMSE
4.9	0.1037 and 0.1311	Networks with low vMSE
4.10	0.1880	Network from early in training training
4.11	0.0815 and 0.0901	ROCs derived from 26 training images.

Table 4.3: Summary of networks and associated vMSE and ROC graphs.

Figure 4.12 shows the intensity distributions of the vessel and non-vessel pixels for all images in the verification dataset. The spatial distribution of these pixels is shown in a few images of the actual neural output; in Figure 4.13 the grey level represents the evidence measure of the processed images. The effects of the thresholding these grey level images is illustrated in Figure 4.15 for the AH verification images and in Figure 4.14 for the RBH verification images.
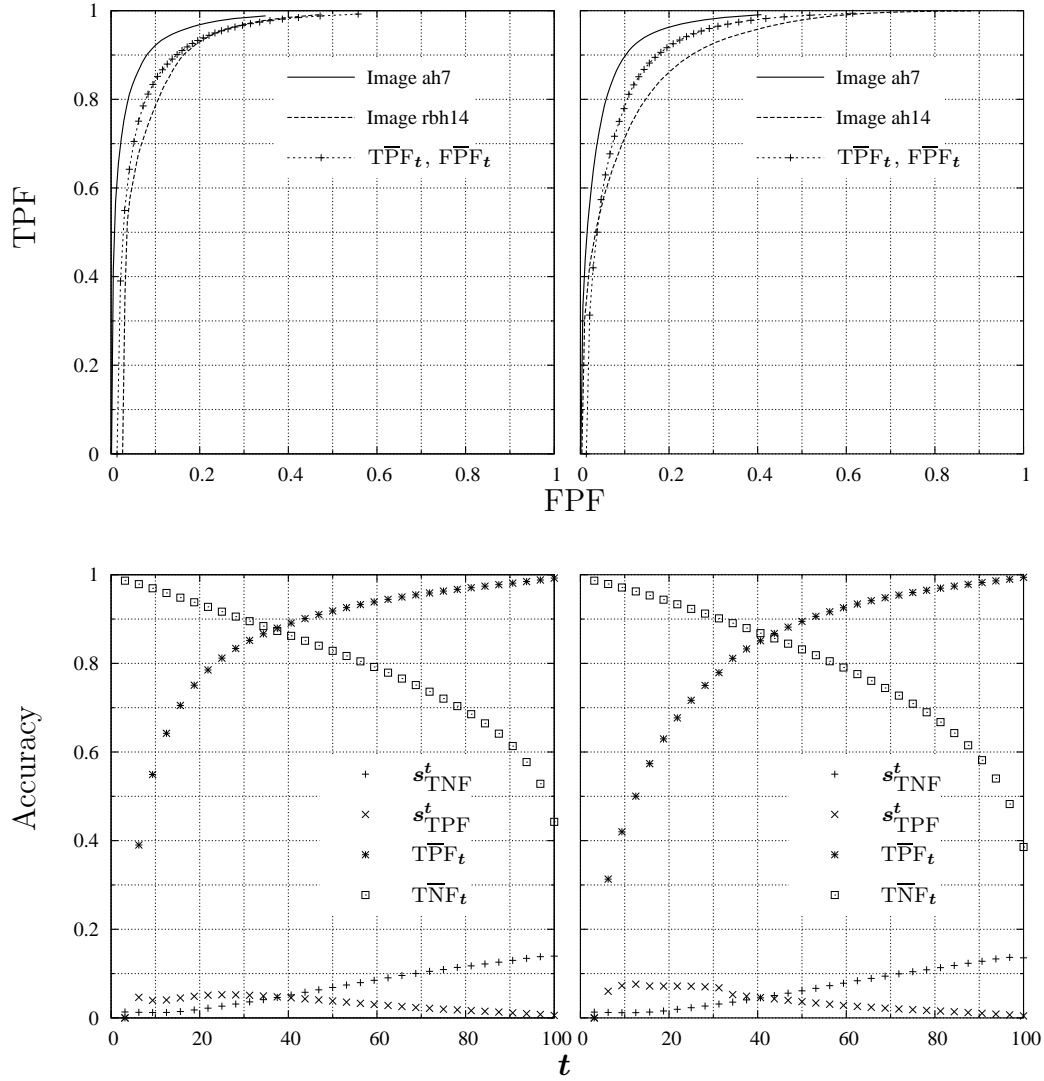
Figure 4.8: Mean, worst and best Reciever Operating Characteristics for a network of vMSE of 0.0815 (top left) and for a network of vMSE of 0.0901 (top right). Lower graphs plot corresponding variance of operating points, as well as the individual specififities and senstivities. (T$\overline{\text{P}}$F$_t$ and T$\overline{\text{N}}$F$_t$ respectively)
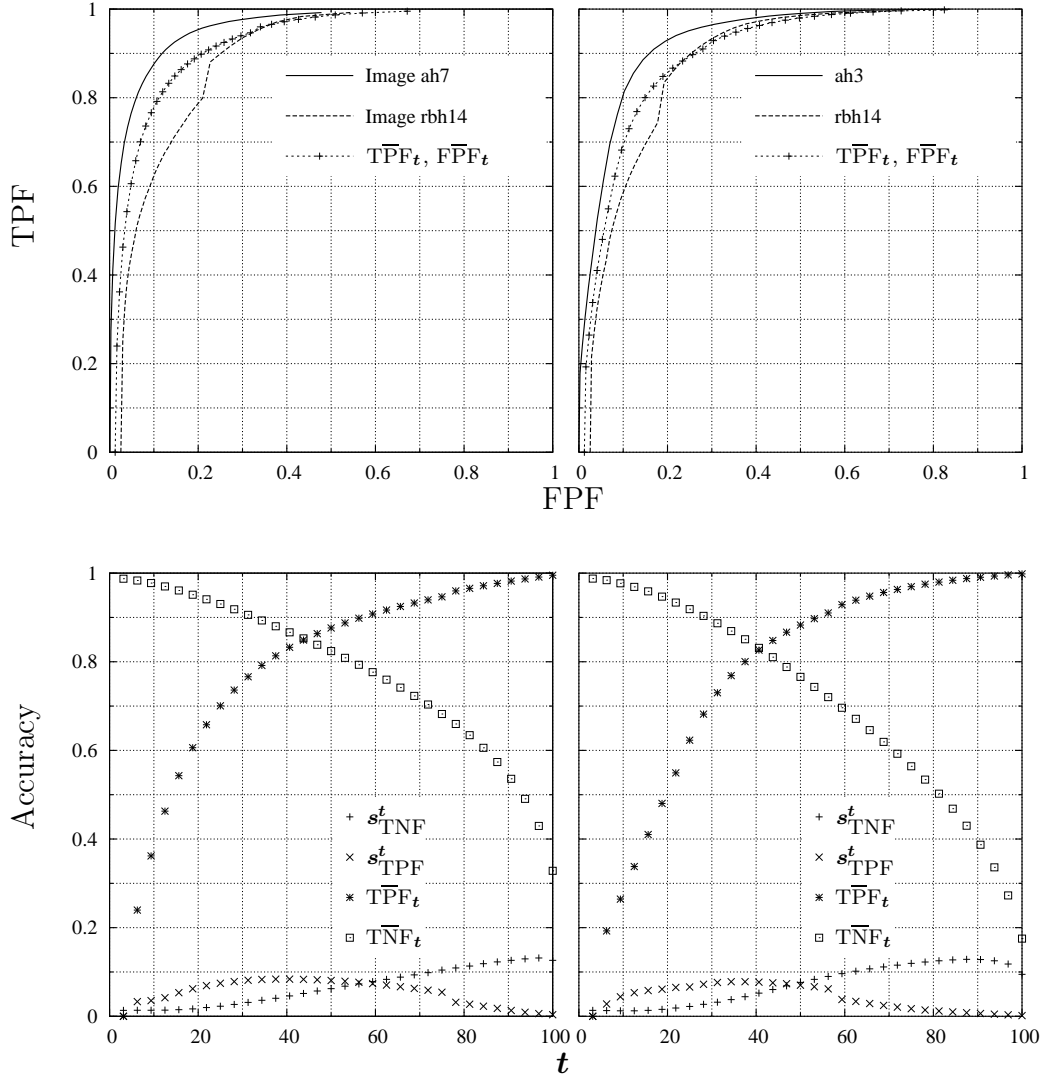
Figure 4.9: Vessel and non-vessel classification accuracies for a network of vMSE of 0.1037 (left) and for a network of vMSE of 0.1310 (right)
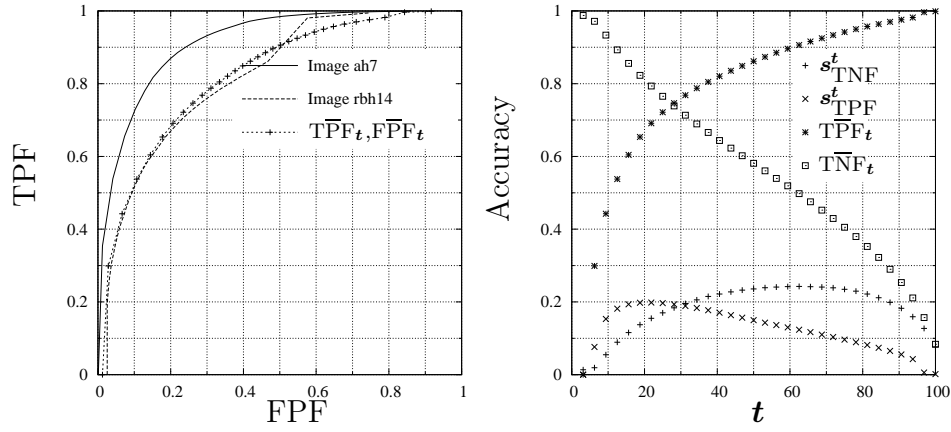
Figure 4.10: Operating Characteristics for a 'poorly' trained network of vMSE=0.18000. Note lower mean TPF and TNF as well as high variability in classification behaviour visible in lower graphs.
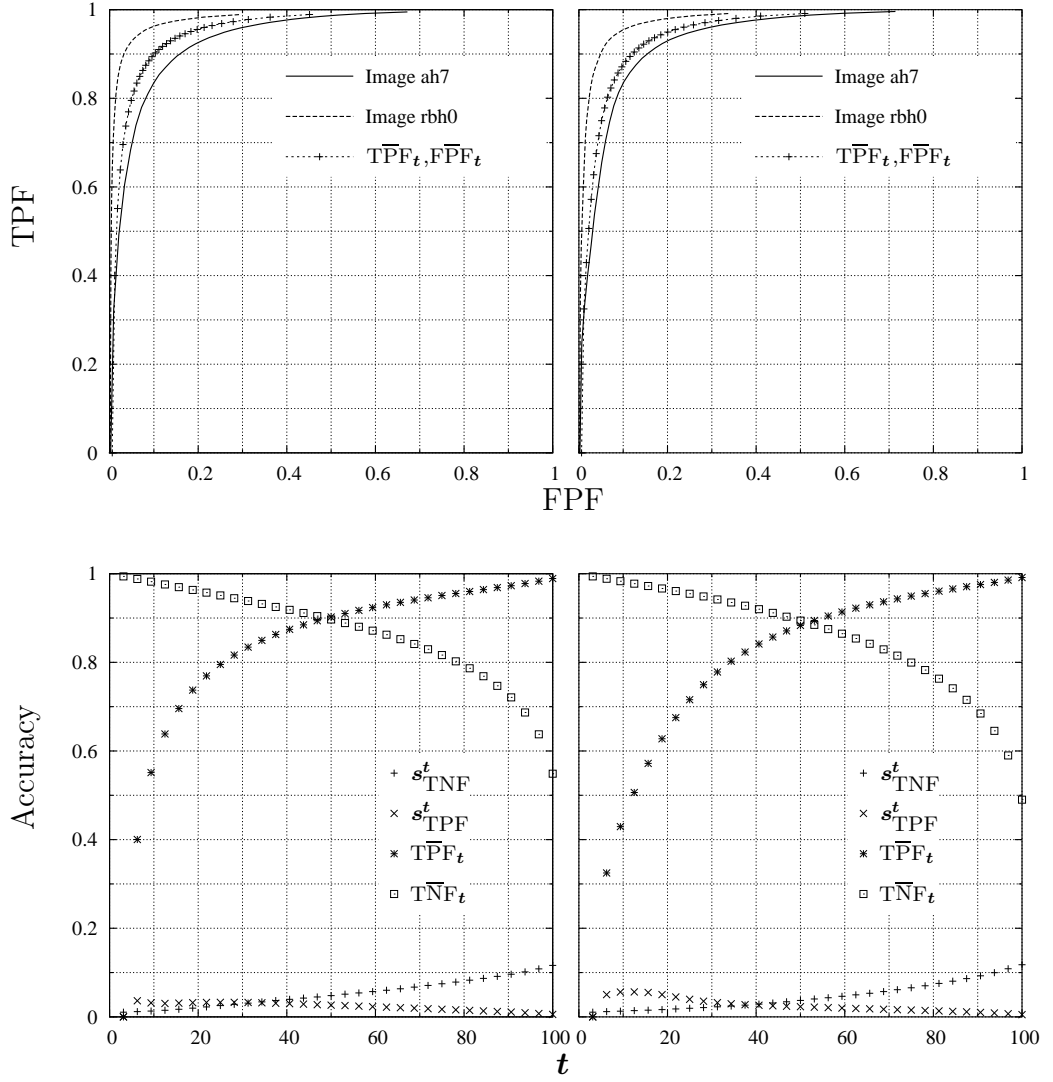
Figure 4.11: Vessel and non-vessel classification accuracies for networks tested on data used in training. These networks had a MSE or 0.0815 and 0.0901.
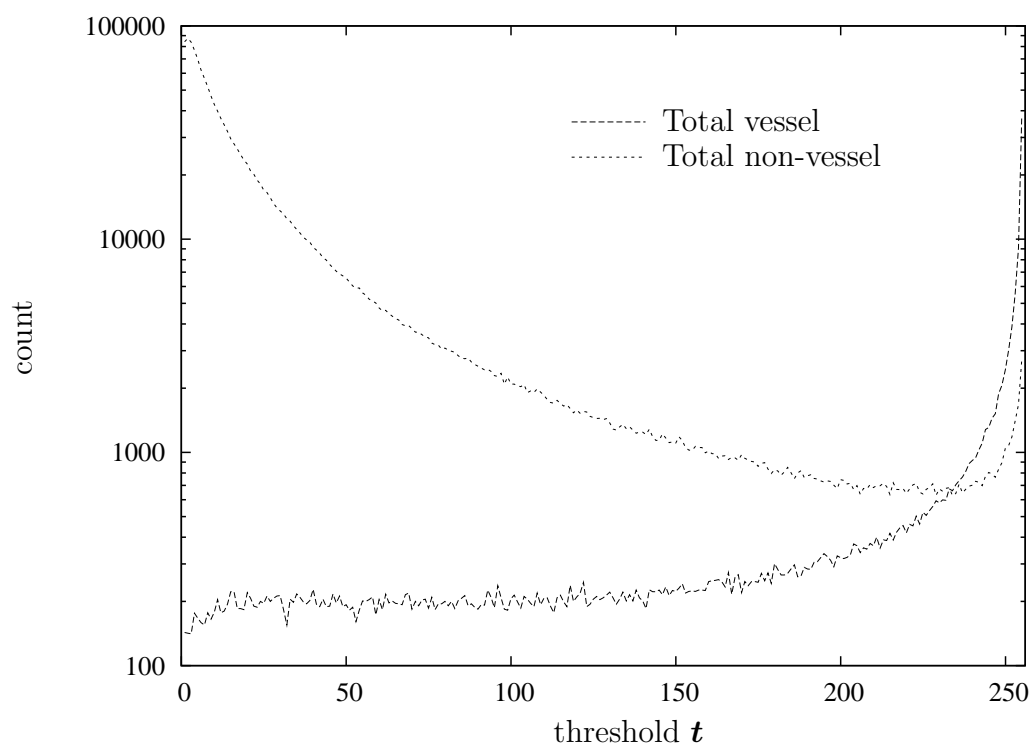
Figure 4.12: Vessel and non-vessel pixel counts for neurally processed AH verification images, dataset.

(a) AH Image 3
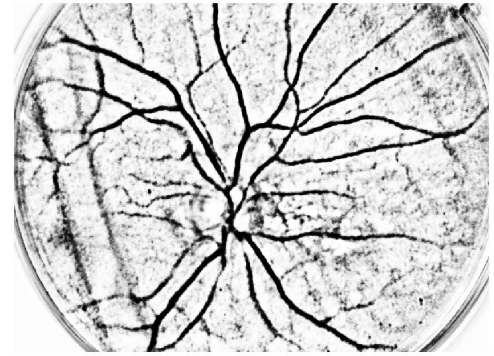
(b) RBH Image 4
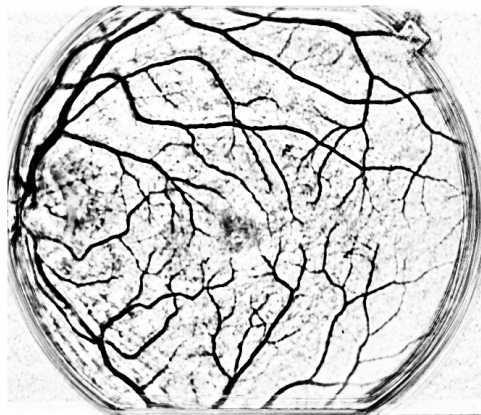
(c) AH Image 18
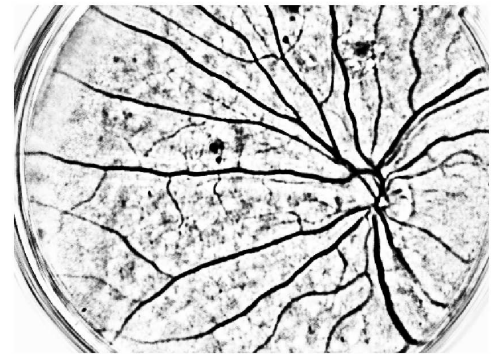
(d) RBH Image 14

Figure 4.13: Verification images processed by network of vMSE=0.0705
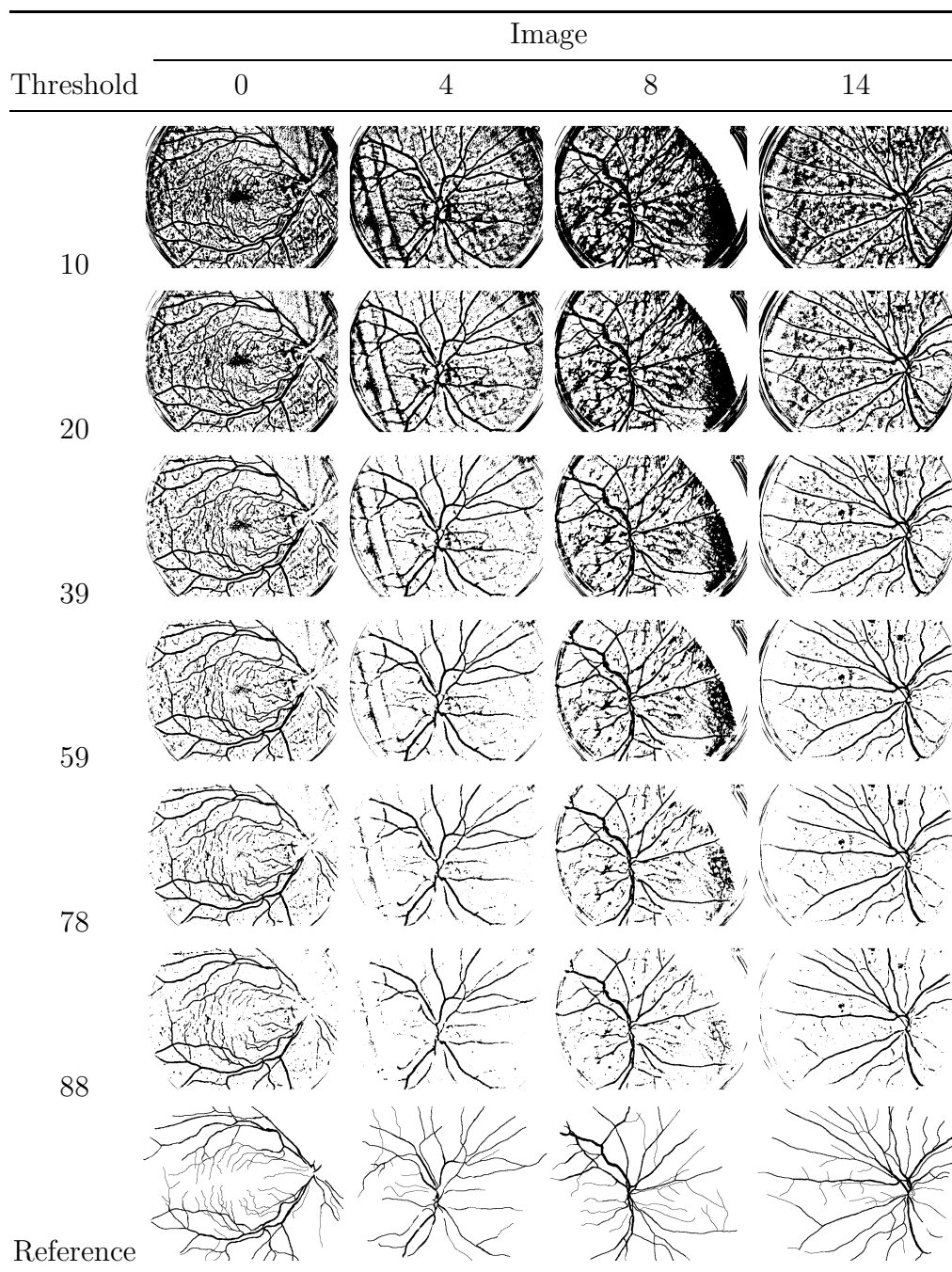
Figure 4.14: Output of the ANN pixel classifier at various values of the threshold parameter for the four images of the RBH images.
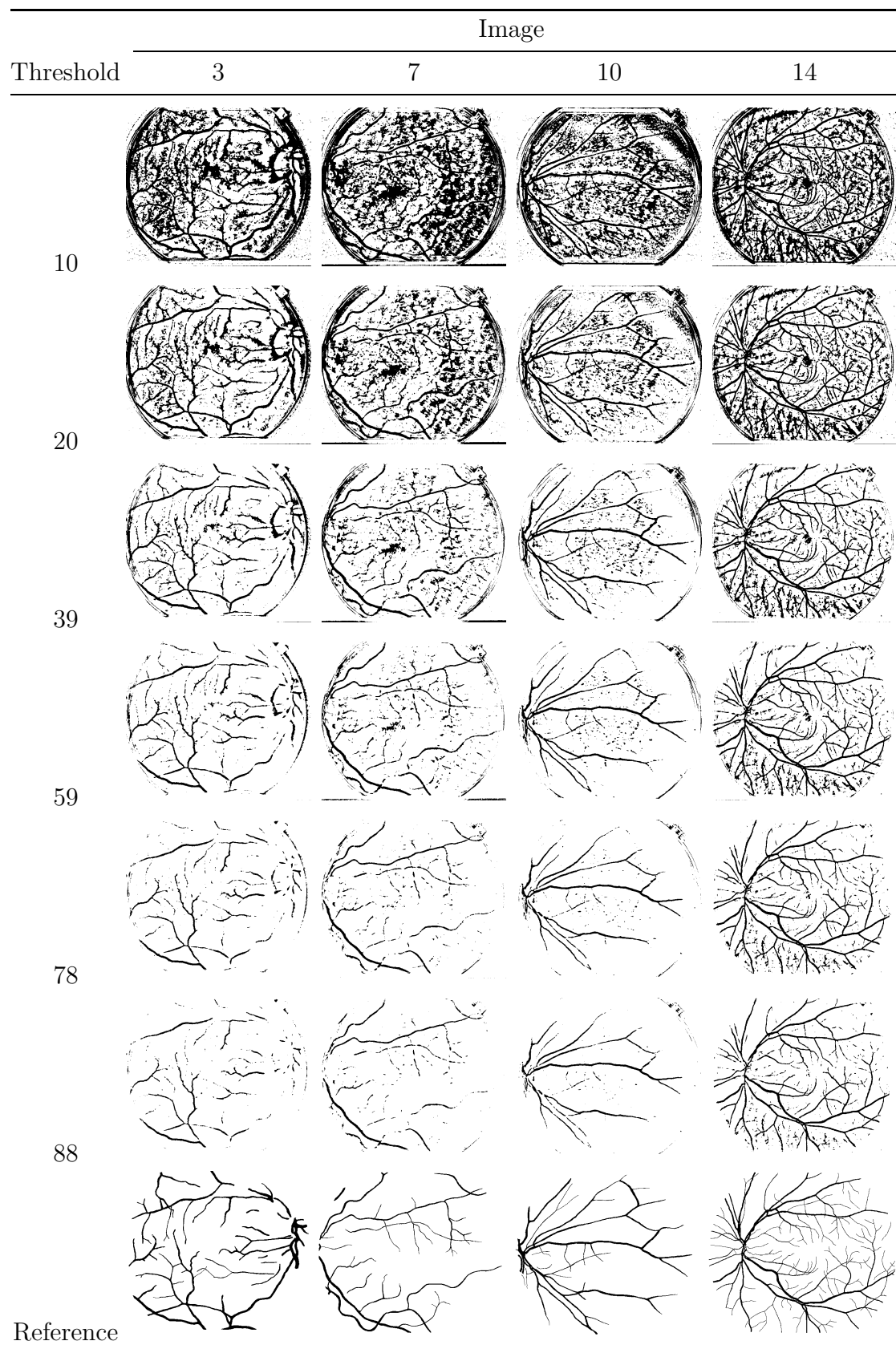
Figure 4.15: Binary classification of pixels for AH images for ANN of lowest vMSE (0.0705).

4.6 Discussion

In this section the raw outputs and ROCs for the various networks are considered. In order to summarise the large quantity of data, points of interest on the ROCs are chosen and the specificity and sensitivity and their associated variability are given. A further measure to summarise ROC is also discussed.

Sub windows of the source images and the processed output images are extracted and used to illustrate the performance in areas of vascular and retinal pathology. In the last subsection the advantages and disadvantages of ANN based classification are highlighted.

4.6.1 Results discussion

The ROC curves and their ancillary measurements for all the ANN networks are now considered. Performance at the TPF=TNF point is compared for various values of vMSE. A discussion on determining the significance, or establishing confidence intervals for the ROCs concludes this subsection.

Unlike the morphological approach, neural performance was not categorised according to the image quality or pathological content. There are two reasons for this. Firstly the number of verification images (9) is small this would lead to single or pairs of images constituting some categories. Secondly the network is trained as far as possible for the current parameters and datasets. This is shown by the decreasing MSE but static vMSE. Had the vMSE not stopped dropping, and no demonstrable difference between verification and training data classification rates been shown, then the inclusion of the training set in analyses could be illustrative of prospective performance.

From the ROC curves detailed Table 4.3 it may be observed that all

classifiers operate above chance level and that the individual image based ROC curves as well as the aggregated ROC curves are consistently smooth with decreasing gradient from 0,0 to 1,1 in TPF:FPF space. Separation of the vessel and non-vessel classes is appreciated in distribution of the ANN outputs when keyed by their class attributes as shown in Figure 4.12.

For the network of lowest vMSE (0.0705) the specificity and sensitivity was observed to average 89%. The classification accuracies for individual images, summarised in the form of the standard deviation (s_{TNF}) is 6%. Similar results are obtained for networks trained down to a vMSE of 0.0815, where the TPF $\approx$ TNF $\approx$ 0.88, and a s_{TNF} of 5% was achieved.

The influence of vMSE on the ROC curves may be observed for changes of approximately 0.01. The shift in the TPF=TNF point is consistent in the direction expected, with higher accuracies at lower vMSE. Despite the noisy verification vMSE over the training epochs, the network at vMSE=0.1880 has the most erratic classification behaviour (Figure 4.10). It has been noted that as the vMSE decreases across the various networks, the s_{TPF} and s_{TNF} also drop at the point where vessel classification accuracy equals that of the non-vessel.

The greatest accuracy and consistency of classification is observed in the ROCs obtained from the 26 training images. Figure 4.11 shows a specificity and sensitivity of approximately 90%. This heightened accuracy was found at an evidence or output level halfway between the ideal targets (i.e. $t = 50$), and is associated with a s_{TPF} and s_{TNF} of less than 5%.

The variability of the individual classifications associated with all aggregate ROCs present consistent trends between the two extremes of the evidence measure. When the t threshold is low, and the image is being mostly classified as background there is lower variability in TNF as there are few pixels changing from vascular to non-vascular classifications. This can be restated as: at low evidence-measure the false positive count is relatively stable. Conversely at the higher thresholds, the true positive fraction

vMSE	t	TPF/TNF	$s_{\text{TPF}}/s_{\text{TNF}}$ %
0.0705	38	0.89	6
0.0815	38	0.88	5
0.0815†	50	0.90	5/3
0.0901†	50	0.89	4/3
0.0910	42	0.87	5
0.1037	44	0.85	9/6
0.1310	41	0.83	9/6
0.1800	28	0.73	14/18

† represents MSE from pattern sets used in training.

Table 4.4: Accuracies and variability of classification over for different vMSEs. Although the $s_{\text{TPF}}/s_{\text{TNF}}$ overlap, they do so for a particular TPF:FPF pair. As will be seen in the next section, when variability data is taken into account over the entire ROC; networks of different vMSE (but with many overlapping operating points) are statistically different.

is less likely to vary, and thus s_{TPF} is reduced. See section 5.5. To summarise all these results an overview of the sensitivities and specificities and their pertinent variances are given in Table 4.4

The most significant problem encountered was the inability to explore the parameter space due to the computational load. For example; no alternative activation functions or weight randomisation strategies were evaluated. The same may be argued for the weight adjustment algorithm itself and the topology of the network. In the same manner the significance of N_w was not assessed.

4.6.1.1 The area under the ROC

The area under the ROC is used, as a single number, to summarise the performance of the ANN classifiers. The area under the ROC curve, or AUC, is useful in a statistical sense, when the data creating the ROCs, in this case the distribution of the evidence measures (ANN outputs) is sim-

ilar between classifiers (Egan, 1975). Thus, even if the data upon which
the classifiers operate is the same, the AUC seems like a number that is
readily, yet incorrectly, compared between classification techniques. It is
calculated here to demonstrate its potential for providing feedback for opti-
misation.

A more transportable alternative to AUCs have been suggested in the
use of confidence bands (Zou et al., 1997; Macskassy et al., 2003). These
can be derived empirically and plotted as contiguous curves or contours of
the ROC. The drawback of these bands is that a graphical appraisal re-
mains necessary, unlike AUCs where a single number can be appropriate
for the comparison of classification accuracy. In the comparison that fol-
low ANN classifiers trained to various vMSEs are assumed, by visual in-
spection, to produce distributions of evidence measures that are similar
for both these classifiers. Figure 4.16 presents the distributions for vascu-
lar and non-vascular pixel intensities of two ANN classifiers. The general
shape of these distributions can be further compared with those in Figure
4.12 (page 134) of the lowest vMSE classifier.

The area under the ROC curve was calculated using the trapeziodal ap-
proximation, which is a non-parametric and conservative estimate. This
estimate is equivalent to the Mann-Whitney or Wilcoxon statistic, and
returns the probability of the correct ordering of a pair of measurements.
The AUC was calculated as follows:

$$\text{AUC} = \frac{1}{\text{TP.TN}} \sum_{v=1}^{\text{TP}} \sum_{nv=1}^{\text{TN}} \boldsymbol{I}(\text{TP}_{\boldsymbol{v}}, \text{TN}_{\boldsymbol{nv}}) \tag{4.2}$$

where TP and TN are the ordered lists of unique vessel and non-vessel pix-
els. $\boldsymbol{I}$ is the function where

$$\boldsymbol{I(x,y)} = \begin{cases} \boldsymbol{1,\ \ x > y} \\ \boldsymbol{0.5, x = y} \\ \boldsymbol{0,}\ \ \text{otherwise} \end{cases} \tag{4.3}$$
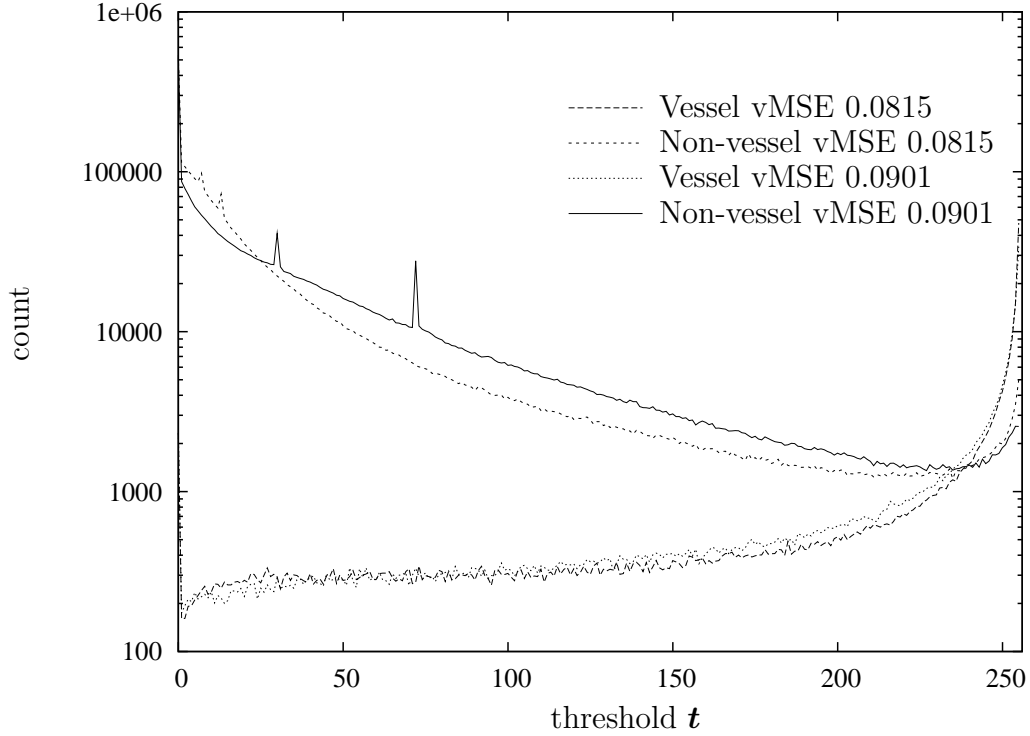
Figure 4.16: Vessel and non-vessel pixel counts for neurally processed verification images, from two classifiers at vMSE 0.0901 and 0.0815.

This holds true whether the pairs represent absolute pixel counts $(\text{TP}_v, \text{TN}_{nv})$ or the their relative fractions (TPF and TNF) as they have been plotted in this thesis. The sampling variability or standard error for the AUC calculated in this manner is estimated using the method of Hanley and McNeil (1992). It is given by :

$$\text{SE}(\boldsymbol{\theta}) = \sqrt{\frac{\boldsymbol{\theta}(1-\boldsymbol{\theta}) + (\text{TP}-1)(Q_1 - \boldsymbol{\theta^2}) + (\text{TN}-1)(Q_2 - \boldsymbol{\theta^2})}{\text{TP.TN}}}$$

(4.4)

$$Q_1 = \frac{\boldsymbol{\theta}}{2-\boldsymbol{\theta}}, \quad Q_2 = \frac{2\boldsymbol{\theta^2}}{1+\boldsymbol{\theta}}$$

(4.5)

where $\boldsymbol{\theta}$ is the AUC. The numbers TP and TN represent the number of vessel (209577) and non-vessel (3136723) cases in the sample. The standard error calculated in this manner can then be used to establish con-

AUC	vMSE	SE(AUC)	± CI (p=0.001)
0.9439	0.0705	3.515×10^{-4}	1.086×10^{-3}
0.9425	0.0815	3.555×10^{-4}	1.099×10^{-3}
0.9297	0.0901	3.893×10^{-4}	1.203×10^{-3}
0.9259	0.1036	3.986×10^{-4}	1.232×10^{-3}
0.9033	0.1311	4.475×10^{-4}	1.383×10^{-3}
0.8354	0.1880	5.520×10^{-4}	1.716×10^{-3}

Table 4.5: Summary of Standard Errors and confidence intervals for the calculated AUCs, at p=0.001, for ANN classifiers at various vMSEs. Note even though many of the operating points overlap (see Table 4.4) and the AUC's differ by only 0.0014, that training for the lowest possible vMSE is statistically worthwhile

fidence intervals or test for statistically significant differences via critical z values taken from the normal distribution. Confidence intervals for the AUC are given in Table 4.5

This thesis did not have specific hypotheses to test, or comparisons to make between classifiers. Never-the-less the confidence intervals for the AUC in table 4.5 show how few samples are needed to estimate the AUC and accompanying ROC, provided pooled datasets are used. This contrasts with the large image to image variability observed when plotting the raw TPF, FPF. These standard errors also suggests that an aggregation of the total, breaching the image to sample correspondence, could undermine verification.

4.6.2   Comments on classification behaviour

From the thresholded and raw output images it may be observed that in genera the larger vessels are reliably segmented. Inspection also shows many of the non-vascular regions, such as that of the mask and retinal parenchyma, to be classified correctly; these patterns are the bulk of the

classificatory load. The majority of instances of false negative classifications are encountered in pixels belonging to the smaller vessels. False positive classifications are more varied in their nature.

Aspects of the fundus that are misclassified as vessel include: confounding pathology (Figure 4.13 (d)), transition zones between mask and fundus, areas inside imaging artifacts (Figure 4.13 (b)) and to some extent the transition between OD and fundus. Images that include a foveal zone, such as RBH image 0 and AH image 7 also elicit a false positive response at low levels of t (see Figures 4.14 and 4.15).

The classification behaviour of the ANN is now presented in a few cases where exudation, haemorrhaging, MA and vascular sheathing are present. The lesions of higher intensity, such as the hard exudation of RBH images 4 and 14, per se don't cause a false positive response. In the samples of 4.17 these exudative lesions seem to be associated with False Positive responses in adjacent regions. False positive responses are associated with the darker lesions such as the blot haemorrhages and microaneurysms of RBH image 3 and the lower AH image 3.

The vascular sheathing of the upper AH image in Figure 4.17 seems to be treated as a smaller vessel and has no deleterious classification consequences. The false positive response for this aspect of the vessel is debatable because of the abrupt change in labeling technique for this segment. A similar situation arises within the circinate exudative ring in RBH image 14. In summary; areas of normal fundus surrounded by pathology that produce vessel like linear structures are prone to misclassification.

The smaller vessels in both sets of images are segmented in a disjoint manner, however the larger images of the AH dataset seem to be classified with greater fidelity. Occurrences where the ANN classifier has labeled vessels that were not marked during ground labelling are also observed as in the blank area at the 2 o'clock position of Image 0 in Figure 4.14.
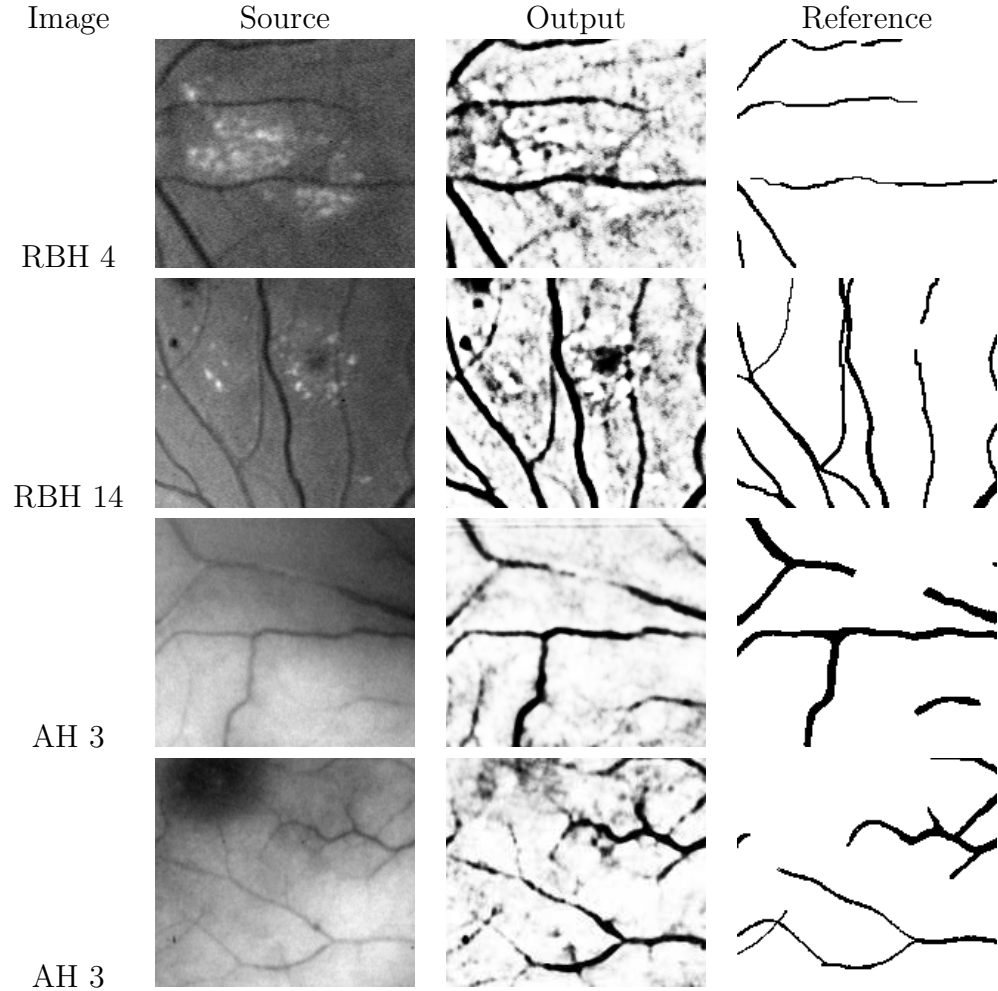
Figure 4.17: Samples of ANN output for areas of pathology.

4.6.3   Ground truth labeling

As discussed in the morphological approach image labeling is a time con-
suming and deliberate process. This is especially true for the smaller, fourth
or fifth order, vessels. These smaller vessels, i.e. those that are a pixel or
two wide, were harder to label and required more corrections. In terms
of their contribution to the pattern counts, it is suspected that they con-
tribute fewer patterns yet are of comparable length (by visual inspection
See Figure 3.3 page 58) to the larger vessels.

The technique, or repeatability, of the human labeling process is of particular relevance to the ANN approach because the data is used to generate (or train) the classifier. Therefore the labeling process must capture the underlying mechanism by which the fundus data is generated. Repeat labeling was found to be 94% sensitive and 90% specific (Hoover et al., 2000). Although the remaining error is not categorised according to vessel size, or characteristics of the surrounding retina. An important assumption is made that the labeling error is random and that the large number of samples that were available and used would nullify any ground truth labeling deviations.

The instances where ANN processing exceeds that of humans are suspected to be rare and difficult to measure. This classification behaviour is the desired goal and the number of false positives that are better reclassified as true positives is not an immediate problem, but should eventually be considered.

4.6.4   Discussion: Artificial neural network methods

The drawbacks in convergence speed associated with standard back-propagation of errors, where training could last up to a week (but typically 3 days) were considered to be offset by the reduced number of parameters and the relatively simple algorithm. The other prerequisites of back propagation ANN were also attractive as they matched the resources available. That is a large multi centre, and multi operator labeled, database. The directly connected or pixel-based ANN bypassed the selection of feature vectors.

Despite the size of the training set the number of vessel patterns was exhausted. A significant proportion of the 597k vascular training patterns represent the interior of the larger vessels. The same is true for the bulk of the non-vessel patterns; as the majority of these arise out 'featureless'

retinal tissue. What is more the network memorised the patterns from the 26 images. A larger dataset would therefore be beneficial.

4.7   Further work

Many aspects of the neural approach undertaken by the author are open to modification and improvement. Most stem from a revised marking or human grading-scheme that would include new label types. The following subsections suggest strategies for the creation of these new data-sub-sets and the subsequent architectures and training of the neural networks that could as a consequence be evaluated.

4.7.1   Reference image labeling

The issues raised in the discussion about ground truth labeling are now addressed. Further work should consider three factors or aspects of labeling when generating more patterns.

First, and most importantly, is the diversity of the patterns. Although this introduces further data categories and more variables to contend with; the patterns should provide adequate representation of all the marginal and infrequent pattern types. The RBH and AH datasets were biased towards images with pathology or quality issues, however the number of patterns they provided for the false positive and false negative cases is felt to be small. In essence a third, non-binary, class-label, capable of directing the random sampling for the creation of the training pattern sets would enhance training.

The second aspect suggested is repeat labeling of the images. The multiple instances of data would enable the use of the agreed subsets to create training patterns of greater consistency. This data could also be used to assess the labeling accuracy of human markers.

Thirdly, consideration should be given to the variety of retinal and retino-vascular presentations. The diversity of healthy and non DR related pathological manifestations is considerable. The present dataset of 35 images does not contain examples of vessel occlusions, congential tortousity, significant pigmentary variations, macro vessels or arterio-venus malformations. Exhaustive vascular identification for DR analysis should behave predictably in the presence of congenital malformations and photocoagulative scars for example.

4.7.2 Modification of the pattern sets

As mentioned in the reference image labelling, partitioning of the training data is a area rich in possibilities. Varying the composition of the training dataset; for example with equal number of samples from each contributing category of false classification (See 4.6.2). Further benefits may be derived from using pattern sets of particular orientation or calibre.

A good examples of mis-classified patterns are the smaller vessels, samples from vessels a single pixel wide or patterns taken from the pixels in transition from vessel to pathologies. Pattern subsets of these cases may improve training, and consequently classification.

A similar idea of separate datasets can also be applied to the non-vascular patterns. Would the inclusion of patterns from pathological areas or their adjacent areas improve training and classification ? In a similar fashion would the increased use of training patterns taken from the transition of the orientation mask to the retina proper, or those taken from optic disk to the retina proper subsequently improve the classification accuracy ?

In addition to varying the distribution of pattern cases, the pattern sets can also be re-sampled. That is varied by selecting a different image sets for training and verification over different networks. Investigating the

classification accuracy by re-sampling the databases using leave-one-out or round robin strategies may also be of interest if resources are available.

4.7.3   Improvements and alternatives to network structure

The structure of the network itself is now considered for further investigation. The datasets outlined in the previous section lend themselves to considering hierarchical networks or sequential training. Future work would also include studying the effect of reducing the input dimensionality and that of the hidden layer.

Hierarchical networks with components trained to recognise distinct features or aspects of the vessels may integrate to obtain better classification accuracy. Emphasis could then be placed on differing networks for the classification of the datasets outlined in the previous section.

The shape of the input sample space is 41 pixels square. Would a rectangular, circular or elliptical sample provide better pattern sets for training and classification ? The potential of scale space techniques is also interesting, research exploring the use of larger and sub-sampled areas of the fundus may be of interest.

4.7.4   Alternative training

Feed forward networks with error back-propagation using the delta rule were chosen for their established nature and the wealth of other research in the area. Further work would include approaches using quickprop weight adjustment for example. This is a variant of the delta rule, where the error surface is assumed to be a locally parabolic, such that larger weight adjustments can be taken. Or the use of a momentum term in the weight adjustment would could overcome local irregularities of the error surface.

If convergence towards a local minimum can be can sped up, from 180 epochs to a more practicable number, then trials assessing the sensitivity of the initial randomisation could be performed. The possibility of pruning of non-contributory connections should also be investigated.

4.8 Conclusion

A neural-network based approach for the segmentation of the retinal vessels has been presented. The range of accuracies with which the ANN classified vessel and non-vessel pixels was given. This range, or tradeoff between the vascular and non-vascular detection ability, was presented using ROC curves. The performance of the ANN can be summarised as achieving an 89% classification accuracy for both vessel and non-vessel picture elements with an image to image standard deviation of 6%.

During training and assessment of the ANN classifier it was necessary that both the set of images, and samples that were derived from them, were chosen in a manner that is not strictly representative of the expected population. In the case of the images; selection was biased so that more than half of the images contained quality and pathological features. These features complicate segmentation and provide a more thorough characterisation of the classification.

The patterns that were generated from these images contained, on average, 6.6% pixels that represented the vasculature. Neural network training did not converge if these class prevalences were respected. Equal proportions of the vessel and non-vessel patterns were therefore used for training and vMSE measurement.

Adopting the convention to quote equal classification rates between vessel and other pixels; the maximum correct classification was observed in a network that was trained down to a vMSE of 0.0704 for a TNF:TNF of

0.89. Further minimisation of the error was not possible as memorisation of the training set was observed.

In summary; this chapter provides, and characterises, the performance of neural based segmentation method for the retinal vasculature. Successful training of a neural-network classifier was achieved using the raw image data. This chapter also shows how a verbal description together with sample output images are complimented by the use of ROC curves. These curves and their associated areas present a summarised description which effectively communicates the performance of the ANN classifier, and to some extent does so, regardless of the implemented instance. However in some situations it was still necessary to clarify classification performance in the pathological or under-represented pixels using examples.

This page intentionally contains only this sentence.

Chapter 5

Comparison and Conclusion

By the end of this last chapter, the important observations from the literature review are emphasized. The neural and morphological segmentation techniques will be summarised further. Then a brief detour is taken to consider these two segmentation-strategies combined to form a hybrid classifier. The performance of all three of these approaches are compared, and to the extent that is possible; placed in the context of previous research.

Vascular segmentation is often used in automating DR screening. This role is further explored, and the consequences of this study are assessed. Finally this chapter and thesis are concluded by presenting the more distinctive observations and results.

5.1   Conclusion from review

Apart from a few profile tracking methods all the vascular segmentation methods reviewed in Chapter 2 have an aim that can be succinctly and unambiguously defined. This aim is to identify vessel and non vessel picture elements. For most of these studies, performance evaluation ends in a pictorial presentation of the 'before and after processing' result images.

In contrast to the bulk of the literature, the author's study is focused on picture elements, somewhat to the exclusion of the higher and more ambiguous concepts of 'vesselness'. This was motivated by the lack of a definition for a vessel, in terms of digital images, that was amenable to quantitive performance analysis.

An important aspect of a method in general lies in the transferability of the methods to datasets outside the study. Peli (1993) highlights these performance concerns when assessing the effectiveness of image enhancements in radiology. Peli goes on to suggest that new techniques should not be compared to un-enhanced or raw images but to a baseline produced by readily available and simple processing methods.

For example when assessing an image enhancement task this should be done against a simple equalisation as opposed to the raw un-enhanced image. Only recently has this type of approach been observed. Use of ROCs was made to objectively demonstrate classified performance (Staal et al., 2002; Jiang and Mojon, 2003). However it is only in the later study that the AH data set was used as a reference. The results of that study will be further discussed in section 5.5.

5.2 Conclusion from morphological segmentation strategies

The morphological approach was capable of segmenting images that were not used during development. The classification accuracy obtained varied, and was observed to be dependent upon the pathology content of the image. This dependency on the presence of pathology was consistent whether quality issues were present or not.

For practical purposes the most important aspect of an actual morphological segmentation is the choice of the t value. The use of ROC curves

somewhat abrogates the selection of this value, but it clearly demonstrates how dependent, the specificity and sensitivity are upon this factor. The selection of a *t* value determines the TPF, TNF trade-off. It also implies the selection of levels of classification variability.

The data presented in this thesis provide the basis upon which to make this choice. A *t* value can be chosen to accommodate, images of different resolutions, pathology content or a posteriori vascular pixels counts. For instance images of smaller fields of view, such as those from the AH dataset, a threshold value of 8 would be more appropriate, to achieve a TPF and TNF of 89%. A higher value of *t*, of approximately 12, would be of benefit for the RBH images to give specificity and sensitivity of 84%.. If mixed resolution images are to be classified a compromise *t* value of 10 would be better.

The classification accuracy (and the image to image variability) expected in a clinical situation are not given because of the biased dataset selection. The reasons for this are further explained in Subsection 5.5.2 where the observed performance bounds are used to estimate expected real world performance.

5.3   Conclusion from neural methods

The small number of images used for the verification of the neural classification does not permit general conclusions to be drawn. The number was not increased because of the labelling effort this would incur. As with morphological segmentation, the neural classifier was capable of classifying patterns not presented to it during development (training). The lowest verification mean squared error obtained was 0.0705 which corresponded to a specificity and sensitivity of 89%. This classification accuracy was observed to vary from image to image with a standard deviation of 6%.

Areas under the receiver operating curves (AUC) were given. The sampling errors associated with these descriptive statistics were small. The statistic was suitable for comparisons of classifiers based on the same data, but is of limited utility to compare classifiers tested on different datasets as differences in classification accuracy between images are masked.

5.4 Discussion of accuracy measurement

This section discusses the utility of the accuracy measurement. The implications for diabetic retinopathy screening are also discussed

If performance characteristics of classifiers are presented without ROCs, comparing retinal vessel and non-vessel classification-accuracy between studies (and in this case methods) is difficult. For example; Hoover et al. (1998) report their sensitivity at 80% and their specificity at 90%, on the other hand Staal et al. (2002) report their sensitivity at 70.0% for a specificity of 92.3%. The questions arises; what was the specificity in the later study at a sensitivity of 80% ? Likewise: what is the sensitivity in the former study at a specificity of 92%. And so on. Clearly two operating points are insufficient to determine, or fail to provide an insight to, the relative performances of the classification methods.

With an ROC, the comparison of classifiers raises many statistical issues. Therefore some studies present aggregated ROCs and leave the data to a visual interpretation with out providing a statistical analysis, see for example (Kanungo et al., 1993; Dougherty et al., 1998; Jiang and Mojon, 2003). The presentation of these curves may be enhanced by providing the variance data as they author has in Section 4.5.

The presentation of results over a range of operating points, via ROCs, provides a more thorough or complete assessment of classifier behaviour. However the effect this has on comparisons between classifiers is complex

and much depends on the ground truth and the datasets from which it is derived. An AUC of a classifier operating upon difficult data maybe equal to that of one operating on trivial data. Even if the same data is used, the area under the ROC curve maybe equal for two classifiers, yet performances may vary at different threshold or cutoff values.

Egan (1975) describes the ROCs that are used for AUC comparison as coming from an 'ordered family of curves'. That is each curve in the family originates from a single function where a parameter, or parameters, have changed. Another way of stating this is that the distributions of the evidence measures must be similar for comparisons of AUC to have statistical validity. This constraint impedes the statistical comparison of ANN derived ROC's and their corresponding AUC with those derived from morphological methods because the distributions presented in Figures 3.35 (page 107) and 4.12 (page 134) are interpreted as being from different distributions.

The accuracy measurement employed in this thesis presents both the aggregated or pooled responses as well as the classification behaviour when the data is organised into images. This makes for more laborious comparisons with other studies, as is shown in the next section, but it doing so the performance is characterised fully.

5.5   Differences and Synergy of the two methods

In this section, classifiers that combine the neural and morphological outputs are presented. The performance characteristics of all three approaches are compared to each other and to those studies that have published accuracies for vascular segmentation.

For the combination of classifiers a variation of the final segmented

morphological images is used. As the last stage of the morphological approach used a logic operation a binary output was produced. Short of a masking, this output is less ameanable to be combined with the output of the neural approach. The hybrid classifier was produced by taking the product of the intermediate-image (after the rolling-ball morphological opening) and the neurally processed image. The ROC for this hybrid classifier is plotted in Figure 5.1. A similar combination of classifiers is produced by the the sum of neural output and the rolling ball intermediate-image, this is presented in Figure 5.2
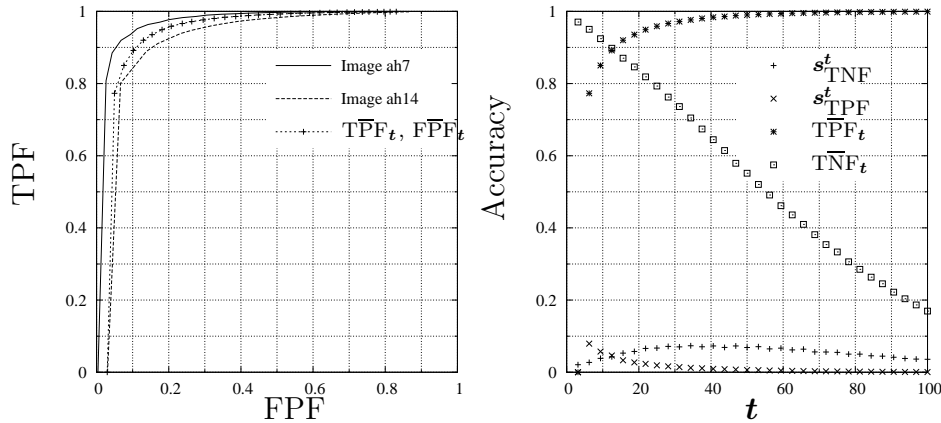
Figure 5.1: Operating Characteristics for the product of the neural output and the rolling ball background subtraction.

Although it makes no physical sense to interpret the sum or product of classifiers as a quantity, the resulting operating points are notable. The segmentation achieved by means of the product of the gray scale morphological opening and the neural output is better than either classification alone. The improvement is derived from the combination of high specificities and sensitivities whilst maintaining a low inter-image classification variability. The equal cost operating points for the hybrid classifiers are summarised along side the results that were available from the literature
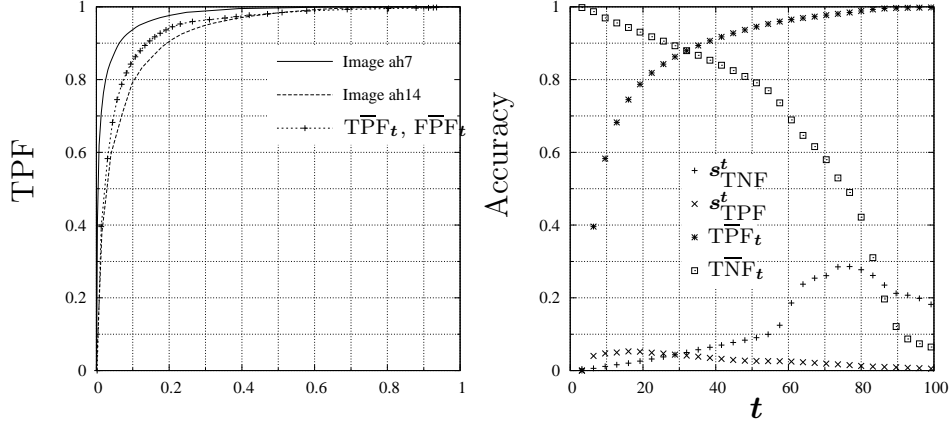
Figure 5.2: Operating Characteristics for the classifier resulting from the sum of the neural output and the rolling ball background subtraction.

in table 5.1. These tabulated results are re-presented in Figure 5.3 using ROCs where possible.

The recent publication by Jiang and Mojon (2003) used the same AH data that the author used in this thesis. The results given are the same both for normal and pathological images. However data in this thesis are accompanied by individual responses, descriptive statistics there-of, and are available over a wider range of operating points.

Although tested on a smaller set of images, the data suggests the neural approach did not suffer from performance degradation resulting from mixed resolution images. In a similar fashion the variability of the neural method is low compared to similar sized test sets from the morphological results. Networks trained to vMSE 0.08 or 0.07 had s_{TPF} that were $\approx \mathbf{0.05}$. In contrast morphological approaches displayed two to three times as much variability durring classification of vessel pixels.

Study			TPF	TNF	s_{TPF}^{t}	s_{TNF}^{t}
Morphological						
	AH data	Combined	89	89	14	7
		Normals	92	92	10	4
		Pathological	87	87	9	9
	RBH data		84	84	16	10
ANN						
	vMSE 0.0704		88	88	6	6
	vMSE 0.0815		87	87	5	5
Combination						
	Product		89	89	5	5
	Sum		88	88	5	5
Sinthanayothin (1999)			83	91	16.8	5.2
Jiang and Mojon (2003)						
	AH data	Combined	89	89	n/a	n/a
		Normals	92	92	n/a	n/a
		Pathological	87	87	n/a	n/a
Hoover et al. (2000)			84	84	n/a	n/a

Table 5.1: Comparison of TPF,TNF for four studies, all units are %. Note results for the Jiang and Mojon study are the same as the author's results.

5.5.1 Performance characterisation

To conclude this subsection on performance evaluation the author proposes that the neural and morphological methods are suitable for the baseline references suggested by Peli (1993). Both approaches are simple, they operate directly upon the pixel data, and robust, have been prospectively tested against a 'difficult' dataset. These two methods, when compared to the alternatives for such a baseline reference; such as a simple threshold or a Gaussian matched filter provide a more challenging and non-trivial approach by which to characterise novel segmentation procedures. The ROC's of Figure 5.4 demonstrate this.
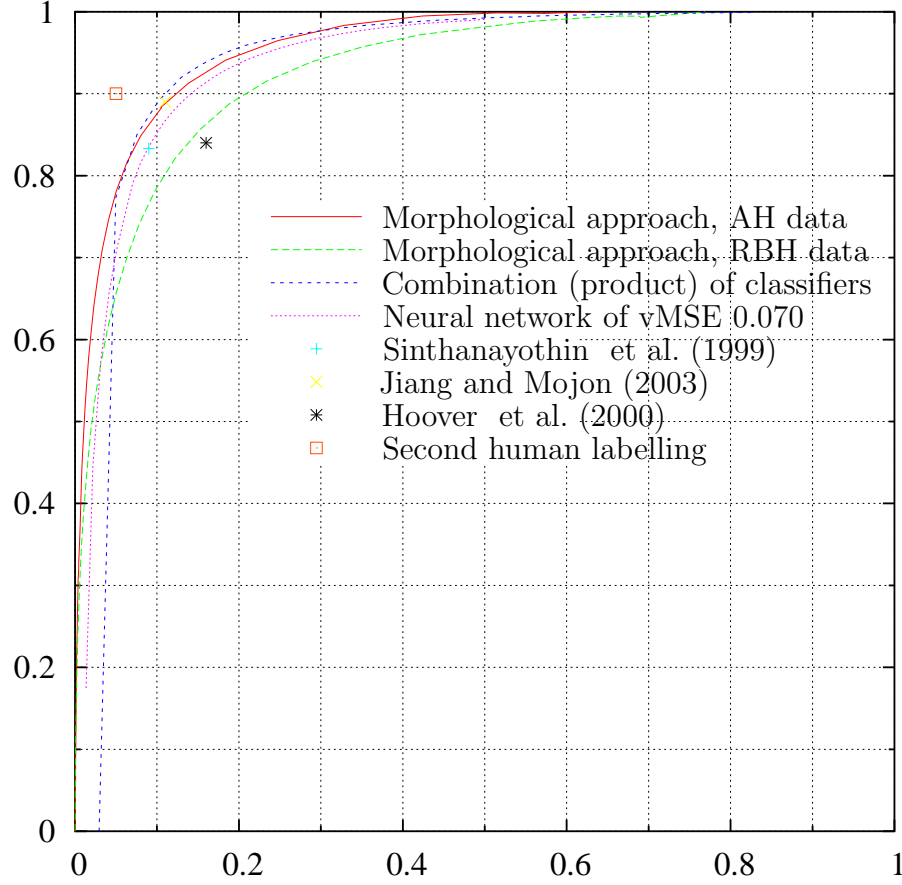
Figure 5.3: ROCs for all classifiers

5.5.2 Vascular segmentation for DR screening

Identifying the vascular pixels in a digital reflectance image of the fundus is considered an important step towards the automated analysis or screening of DR. Once the vessels have been identified analysis of the parenchyma can proceed. The lesions, colour and texture of the space inside the vascular arbour can now be directed according to its distance from the optic disk and nearest known vessel.

The retinal vasculature has a rich and varied set of healthy and pathological features. In health features such as the changes in calibre, distri-
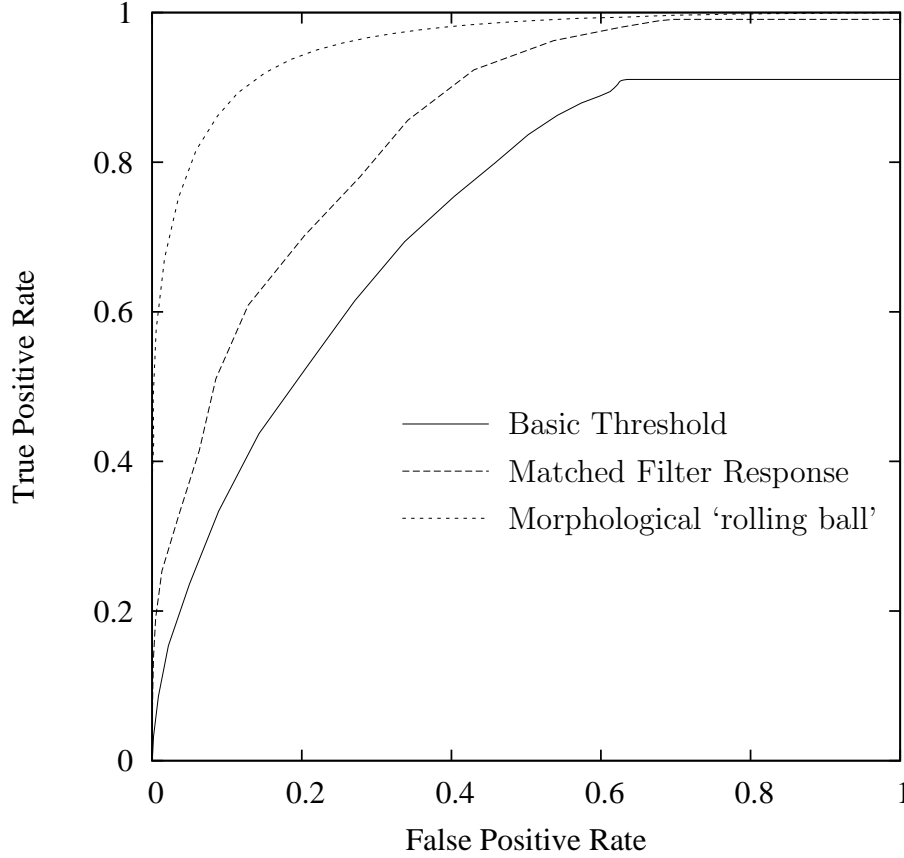
Figure 5.4: Comparison of three 'simple' vascular segmentation approaches. the top trace of the morphological rolling ball may be suitable as a baseline to which other, more sophisticated, strategies could be compared. The Simple threshold and the Matched Filter Response (Chaudhuri et al., 1989b) discussed in Section 2.6.3 do not segment reliably enough.

bution of capillaries and diverse pigmentation are compounded by normal variations acquired during image acquisition. This thesis has taken these presentations into consideration. Lesions such as exudates, sheathing, mild, moderate and severe haemorrhaging, cotton wool spots and ischaemia have all been present in the accuracy assessment. However these complications are present to such an extent in the datasets that they are no longer repre-

sentative of the expected prevalences in a screening scenario.

The experimental design chosen by the author trades the measurement of vascular segmentation accuracy of clinically representative data, for the estimation of its lower and upper bounds. If the vascular segmentation strategies evaluated in this thesis would be applied or incorporated into an automated screening system, then the accuracies presented need to be adjusted to reflect the contents of the dataset selection. In the case of the morphological approach the upper and lower bounds are a specificity and sensitivity of 92% in the best case scenario; i.e. all images normal, of good quality and higher resolution and in a worst-case scenario 82% for mixed resolution images with simultaneous pathology and quality issues.

Thus with the substantial majority of images being free from quality or pathological complications the TPF and TNF as well as their associated variances can be expected to shift towards the normal good-quality performance curve.

5.6   Assumptions made in this thesis

There are two assumptions which would benefit from further scrutiny. First there is the assumption that human graders are capable of labeling the retinal vessels in a consistent and accurate manner. It could be argued that the labelling is subjective, that the vessel edges are hard to establish, and that the binary representation of a vessel edge is inadequate. However the assumption may be justified by using the vascular classification of a human in the subsequent analyses of any method dependent upon vascular segmentation.

The second assumption is more fundamental and is part of the feature selection problem. Given the segmented vessels are to form part of a greater system with an end point other than the identification of the vas-

cular pixels themselves. In this situation the intelligent system may be argued to emulate the established analytical process of the retinal specialist. This mimicry of the diagnostic protocols could be warranted if intermediate steps, are necessary to justify an diagnosis.

However with gold-standard diagnostic labels a more abstracted or direct analysis could be attempted of the retinal image or the segmented vasculature. These alternate analyses could include texture, colour, perimeter to area indices, of the vessel image for example. In this way Baxt (1994) and Kristinsson et al. (1997) found novel indicators of disease in myocardial infarction and macular oedema respectively. Conversely these factors may be used to reliably identify normal images so as to reduce the interpretative load of DR screeners.

5.7 Further work

There are many parameters which were not fully investigated, and many more which can be created. On the neural approach, activation function, weight pruning and the initial randomisation are left unexplored. For the morphological approach pre or post-processing steps, the effective of connected component removal or automated masking could be investigated. However the two factors that are expected to bear the greatest impact are the ball size in the morphological approach or input dimensionality of the neural network.

Improvements to the image dataset are also positively considered, both in terms of size, diversity and representativeness. Although some of the images were of very poor quality, no images with extremely severe compounding pathology were included. Examples of these types of images include colobomas, extensive vitreous haemorrhages, severe photocoagulative scars and pigmentary lesions. Another type of retina which was not present in

the data is the apigmented retinae. There are also varied, rare and peculiar vascular anatomies such as anastomoses and congenital malformations. For completeness these compounding pathologies should be included into the image datasets. The classification for normal, pathological and unusual retinal states should ideally be the by product of a large randomised trial, where a standardised grading protocol has been employed. Good examples of this are the ETDRS and the DRS.

The creation of the second class label as suggested in section 4.7.1 would also be best served by persons with extensive interpretative experience. These markings need not be exact as the vascular ones in many cases, as they mostly serve to control the distribution of pattern classes. The benefits of this would be twofold. Such patterns would delineate the boundaries between classes in pattern space, and would also provide feedback for methods not based directly upon these input patterns.

5.8   Final Summary and Conclusion

This thesis evaluates the performance of two vascular segmentation strategies. The results are unambiguously presented with full details on the image dataset used and the variances observed. Degradation of segmentation accuracies were observed for pathological and poor quality images. These differences were not measured using the neural network approach. However the neural verification dataset did include four images with pathology and two of poor quality, and produced the lowest variances approximately of 6% for a sensitivity and specificity of 0.88. These accuracies were improved upon by combining the output of both segmentation strategies. However the observed improvement was small.

From the morphological approach we conclude that the presence of pathology has a negative impact upon classification accuracy. The neural

and hybrid approaches seem to have higher specificities and sensitivities which are also more consistent; i.e. they have smaller image to image variability but the evidence is less conclusive.

As an objective measure of vascular classification accuracy, per-pixel sensitivities and specificities raise many questions. If the concept and or definition of a retinal vessel can be made then its subsequent identification and segmentation would be easier. However, currently it was found that the most amenable and objective measure was provided at the pixel level.

Using these pixel level statistics, this study performed the widest objective and prospective assessments known to the author. In these assessments vascular and non-vascular classification accuracies were demonstrated to be similar and more consistent to those published in the literature. What is more these performance characteristics were obtained from a comprehensive image set that had been withheld for this prospective verification purpose. This validation, although costly in terms of data, prevents overfitting and goes some way to ensure the applicability of the results to a wider set of images.

Appendix A

Previous Review

Review

Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy

T. Teng   M. Lefley   D. Claremont

Academic Biomedical Engineering Research Group, School of Design, Engineering & Computing, Bournemouth University, Dorset, UK

Abstract—*Patients with diabetes require annual screening for effective timing of sight-saving treatment. However, the lack of screening and the shortage of ophthalmologists limit the ocular health care available. This is stimulating research into automated analysis of the reflectance images of the ocular fundus. Publications applicable to the automated screening of diabetic retinopathy are summarised. The review has been structured to mimic some of the processes that an ophthalmologist performs when examining the retina. Thus image processing tasks, such as vessel and lesion location, are reviewed before any intelligent or automated systems. Most research has been undertaken in identification of the retinal vasculature and analysis of early pathological changes. Progress has been made in the identification of the retinal vasculature and the more common pathological features, such as small aneurysms and exudates. Ancillary research into image preprocessing has also been identified. In summary, the advent of digital data sets has made image analysis more accessible, although questions regarding the assessment of individual algorithms and whole systems are only just being addressed.*

Keywords—*Diabetic retinopathy, Image processing, Screening, Automation, Intelligent systems, Fundus image analysis*

1 Introduction

OBJECTIVE DIAGNOSTIC indices in diabetic retinal health care have been studied since 1969 (SKOVBORG *et al.*, 1969). This review summarises the published literature on image processing and automated systems for the screening and diagnosis of diabetic retinopathy (DR). It focuses on image processing algorithms for the segmentation (i.e. the identification) of the retinal anatomy and the lesions induced by diabetes mellitus.

The medical, social and financial rationale supporting regular and frequent screening for the lesions of DR is strong. Detailed and well-documented diagnostic protocols have been developed over the last 20 years (DRS RESEARCH GROUP, 1981; ETDRS RESEARCH GROUP, 1991*a*). It is important to note that timely treatment for DR prevents severe visual loss in over 50% of eyes (ETDRS RESEARCH GROUP, 1985; 1991*b*; DRS RESEARCH GROUP, 1987).

Computer simulations of recommended eye care (AMERICAN DIABETES ASSOCIATION, 1999), performed by JAVITT *et al.*

(1994) and DASBACH *et al.* (1991), demonstrated that prevention and treatment were less expensive than the health care and rehabilitation costs incurred by visual loss or blindness. In spite of this evidence, there is still a need for structured care to impart these potential benefits in the United Kingdom (DIABETES UK, 2000; NATIONAL SCREENING COMMITTEE, 2000). A comprehensive medical overview of DR, including the data required for optimum care, can be found in AIELLO *et al.* (1998*a*).

Current retinal image databases are not sufficiently integrated with hospital information systems to allow correlations between the patient's condition and novel features of the retinal images to be investigated. Instead, research locating or measuring individual anatomical landmarks or specific diabetic pathology is receiving the most attention. Previous papers that review the image processing of retinal image data exist (GOLDBAUM *et al.*, 1990*a*; WARD *et al.*, 1988), but these need updating to include a decade of research. It should also be noted that many studies lack objective results and so cannot be quantitatively evaluated.

2 Image processing for segmentation of the normal retina

From colour transparencies and digital images of the retina, the optic disk, vessels and retinal parenchyma are easily recognised (see Fig. 1). These landmarks give a framework
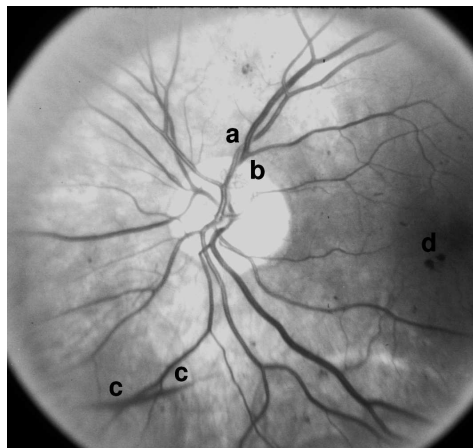
Fig. 1 *Standard image 10A from ETDRS grading scheme. Note variability of contrast due to retinal pigment and imaging. Optic disk (bright circular feature in centre) has (a) twisted artery and vein at its rim and (b) small tuft of neovascularisation within it. (c) Preretinal haemorrhage and (d) intra-retinal haemorrhages are also apparent*

from which automated analysis and human interpretation of the retina proceed. However, the image must be of adequate quality and resolution.

2.1 Image resolution

A monochrome digital image can be represented as a two-dimensional array, with each cell or picture element (pixel) holding a value representing the intensity at that point in the image. This intensity is also called the grey level and is usually limited in resolution by the number of bits (or bytes) that represent it.

The spatial resolution of a digital image is given by its area in pixels, their size and the field of view. For a fundus image recorded on photographic film and through clear ocular media, the resolving power of the whole system is limited by the film and its development. The spatial resolution of typical transparency film used for imaging DR ranges from 1 to 100 cycles mm^{-1} for a response above 10%.*

Film resolutions of approximately 4000 $pixel^2$ have been quoted by SCHALKOFF (1989). This figure is of the same order of magnitude as the 50–130 lines mm^{-1}, or 4600 lines per 35 mm, quoted by the Agfa datasheets. Film-based evaluation of DR uses a relatively small or narrow field of view. For 35 mm film-based stereoscopic evaluation, two 30° field of view images are required. Lesions of the retinal capillaries, such as micro-aneurysms and intra-retinal microvascular abnormalities (IRMA) require this resolution for unequivocal assessment.

Digital image acquisition is usually over a larger field of view; with 45, 50 or 60° being typical. Images are usually rectangular and have the long dimension ranging from 600 to 1024 pixels. Digital images thus represent larger fundus areas at lower resolutions. The film-based reference imaging protocol also uses stereoscopic pairs that the interpreter could integrate to advantage. Digital transducers have linearity and sensitivity advantages over photographic film. The ease and fidelity of

---

*Data obtained from Kodak, Agfa and Fujifilm websites for Ekta-chrome, Agfachrome and Fujichrome transparency films

reproduction and transmission mean that digital imaging will become an attractive alternative to 35 mm film-based methods, once the resolution approaches that of traditional film.

2.2 Image preprocessing

Patient movement, poor focus, bad positioning, reflections, disease opacity or inadequate illumination can cause a significant proportion of images to be of such poor quality as to interfere with analysis. The sphericity of the eye, a significant determinant in the intensity of reflections from the retinal tissues, compounds artifacts such as circular and crescent-shaped low-frequency contrast and intensity changes arising from the interface between the anterior and posterior ocular chambers.

In approximately 10% of retinal images, artifacts are significant enough to impede human grading (KRISTINSSON *et al.*, 1997; KLEIN *et al.*, 1989; LIESENFELD *et al.*, 2000). A similar proportion is presumed to be of inadequate quality for automated analysis. Preprocessing of the fundus data either removes or flags the aforementioned interferences. LEE and WANG (1999) proposed that, by correlating the histogram of intensities of an image with that of an ideal histogram representing the best images, an index describing the contrast, brightness and signal-to-noise ratio of an image could be obtained. The authors describe their results as being in agreement with human perceptions.

OKAZAKI and TAMURA (1983) approximated unequal retinal illumination as spherical. Their surface fitting algorithm was successfully applied to the image after a range of thresholds had been set. The average centre of these binary images provided enough terms to apply a least squares method to find the radius and vertical offset. LEE (1992) found median filtering to be preferable to the fitting of high-degree polynomials, reasoning against non-linear computations and the parameter selection involved. LEE (1992), ØIEN and OSNES (1995), MENDONÇA *et al.* (1999), SHIN *et al.* (1999) and EGE *et al.* (2000) used a 31 $pixel^2$ median filter as a smoothing operator to remove the effect of the background. Other background changes introduced during film digitisation have been removed by subtracting (FRAME *et al.*, 1997*a;b*) or dividing (CREE *et al.*, 1997) the uneven illumination known as the flood image.

Some information about retinal images is commonly discarded before image processing. During colour digital image acquisition, the intensities of the reflected illumination are recorded at three wavelengths. These three wavelengths are chosen to represent the red, green and blue (RGB) spectral ranges of the human photosensitive pigments. Sometimes, only the the green band is used in the image processing of digital retinal images (CHAUDHURI *et al.*, 1989*a*; GOLDBAUM *et al.*, 1989; KOCHNER *et al.*, 1998; MARTÍNEZ-PÉREZ *et al.*, 1999*a*; LEE *et al.*, 1999; ZAHLMANN *et al.*, 2000).

The spectral content of each pixel can be decoupled from its intensity measure, and, with this approach, it is possible to operate independently on either brightness values or its perceived colour. SINTHANAYOTHIN *et al.* (1999) used the hue, intensity, saturation colour space described by GONZALEZ and WOODS (1992) to perform a local sigmoidal histogram stretch on the intensity plane. This increased the local contrast and minimised contrast variations over the whole image. Alternatively, shading and illumination anomalies can be ignored in small regions of interest or sub-windows (AKITA and KUGA, 1979; GREGSON *et al.*, 1995).

2.3 Locating and segmenting the optic disk

The optic disk (OD) is the brightest feature of the healthy fundus; it is approximately circular and measures $\approx 1800 \mu m$ in diameter; the retinal vasculature originates from its centre (see

Fig. 1). The OD lacks a visible capillary supply and appears paler owing to the lack of an underlying pigmented layer.

The contrast between the emerging translucent, blood-filled vessels and the white background was exploited by SINTHANAYOTHIN *et al.* (1999). They used a local measure of pixel intensity variance, of window size equal to that of the OD (80 pixels), and swept it over the image. The measured peak variance was successful in locating the OD in 111 out of 112 test images.

Instead of the variance, the average brightness value alone has also been used to locate the OD (LEE *et al.*, 1999; TAMURA *et al.*, 1988). A combination of many features, such as edge concentration, colour and brightness, was weighted by KATZ *et al.* (1988) to highlight regions with OD-like properties. CHAUDHURI *et al.* (1989b) used local measures of vessel convergence and orientation combined with a matched filter for bright circular objects to locate the optic disk. This resulted in OD localisation of 31 optic disks in 35 predominantly pathological images.

The convergence of the retinal vessels upon the OD was more exhaustively exploited by HOOVER and GOLDBAUM (1998a) and AKITA and KUGA (1982). The latter study identified the order in which the vessels branch from the OD origin and measured the intensity of the fundus around this ranked list. The brightest pixel in the neighbourhood of the highest ranked, or root, vessel was then identified as the OD centre. HOOVER and GOLDBAUM (1998a) used elongated and smeared versions of vessel segements. The now diffuse vessels accumulated and overlapped in areas of convergence, such as the optic disk.

The Hough transform, a method for locating basic shapes such as circles and lines, was used by KOCHNER *et al.* (1998) to locate the optic disk. They segmented the two main vascular arches to obtain points to which an ellipse was fitted. At the apices of this ellipse, the OD was prospectively located using matched filters (see Section 2.4.2) to find the major vessels leaving the OD. The response of the matched filters was maximum if radially distributed vessels were present within a bow-tie shaped region of interest (ROI), centred on the presumed location of the optic disk. A Hough transform for circular feature detection has also been employed (YULONG and DINGRU, 1990). The edge points were derived from the whole image using a series of transformations aimed at detecting the bright optic disk.

WANG *et al.* (1990) fitted circles to the OD by successive approximation. Their approach started with an unspecified thresholding. Pixels were iteratively rejected from the resulting binary image by the discarding of 15% of pixels not within a shrinking circle centred on the mean location. If, after shrinking, the origin of the circle could not be moved so as to reject a further 15% of pixels not within the circle, the OD was taken to be identified.

2.4 Segmentation of the retinal vasculature

This Section summarises research into segmenting or deriving a binary image representing the retinal vasculature. The approaches are varied, and the *a priori* knowledge is substantial, and yet the results are scarce. The vascular tree needs to be segmented so that its physical dimensions can be measured. Length, diameter and path, and the changes induced by the progression of DR, can become valuable diagnostic indices of the disease. Vascular segmentation is also required to stop the curvilinear shapes from interfering with subsequent analysis of DR lesions.

Most of the results published for vascular segmentation are described as adequate or presented as side-by-side comparisons. An algorithm should be assessed against a ground truth or a gold standard, providing an objective and transferable measure for the performance or efficacy of the approach. TOLIAS and PANAS (1998), HOOVER *et al.* (1998b; 2000) and SINTHANAYOTHIN

et al. (1999) have produced the only studies, out of approximately 23 papers on vascular segmentation, to provide results describing the performance of the vessel segmentation algorithms.

Within these three studies, the definition of a true match varies. An exact match between pixels, corresponding to the same location in both images, was required by SINTHANAYOTHIN *et al.* (1999) and HOOVER *et al.* (2000). HOOVER *et al.* (1998b) only required neighbouring pixels to match. Vessel segments, defined as the tract of vessel between branching points, were used by TOLIAS and PANAS (1998). However, in this latter study, it was not clear how the correspondence between human- and algorithm-marked vessel segments was established.

The physical level or *a priori* knowledge of the retinal vessels is summarised in Table 1. Tracking or following the vascular arbour, its branches and convergences exploits *a priori* knowledge of the retinal vascular network.

Tracking saves computation time by avoiding scanning the entire image, whereas matched filters are swept over its entirety. In a similar way, neural networks are scanned over the whole image, the response of the filter or network being proportional to the likelihood of a vessel being present at one of the pixels of interest, usually the centre pixel.

For the purpose of this review, vascular segmentation is split into four separate areas: local operators, matched filters, vessel tracking and neural networks.

2.4.1 *Local vascular segmentation:* The retinal vessels are darker than their surroundings, enabling segmentation of small regions by the use of a threshold. Pixel intensities above or below this threshold can be accepted as potentially belonging to a vessel, or discarded as background. Alternatively, pixel brightness values in a region can be homogenised. Areas of pixels can be clustered into groups, depending on their brightness and how they are connected to each other. This connectivity constraint determines the shape of the area and is usually called the structuring element. Grey-scale

Table 1   A priori knowledge of retinal vascular tree

Vessel	Feature	Description and reference
Segment	calibres	are invariant along their segment length; profiles are detailed by HAMMER *et al.* (1997) and RASSAM *et al.* (1994)
	curvature	vessel centrelines are regular, i.e. contain no discontinuities and have a finite range of rate of change of direction (HART *et al.*, 1997)
	length	is theorised to scale in a fixed ratio (WEST *et al.*, 1997)
	overlap	veins (arteries) do not overlap veins (arteries)
Branch	angle and number	branches from larger vessels to smaller ones are dichotomous and acute
	calibres	calibres of branches cannot be assumed to preserve volume, as blood slows down for metabolite transport
Special cases	twisted pairs	a rare helical arrangement of veins and arteries
	parenchymal dips	parts of segments are submerged within the opaque retina
	cilioretinal arteries	are small, arise from edge of OD and do not form part of whole vascular arbour

structuring elements have an intensity value that is added to the underlying image. If the result of the summation is larger than either the structuring element or the image, then it is kept. This process is repeated for all positions of the structuring element over the underlying image and is called dilation. If the structuring element is subtracted and the minimum value is kept, the process is called erosion. Noise and small artifacts can be removed in this way. Careful selection of regions enables local thresholding to work.

Studies assessing venous beading (KOZOUSEK *et al.*, 1992; GREGSON *et al.*, 1995) measure the variation in vascular calibre along a segment. Venous beading is assessed only in the larger vessels, where pixel intensity distributions are well separated from those of the retinal parenchyma. GREGSON *et al.* (1995) chose the regional threshold as equidistant from the means of background and vessel pixels, which in turn were defined as top and bottom cumulative percentiles; the percentage was not given. Erosion and dilation with a square-shaped operator after thresholding fill in gaps and remove bleeders. Precise estimation of the vessel edge is obtained by dividing the area under a vessel cross-section by the range of heights in its profile.

Some regions of the ocular fundus are devoid of major vessels, and methods, such as that of GREGSON *et al.* (1995), that use local dynamic thresholding fail and produce artifacts. AKITA and KUGA (1979) overcame this limitation by analysing groups of connected pixels. Multiple thresholds of the derivative image generate blobs, allowing area-based discrimination to discard components too small or large to be vessels.

YE and ZHENG (1995) used a spherical structuring element to enhance the difference between the fundus and background. To this image, thresholds over the full range of intensities were applied using the Otsu method (OTSU, 1979), where the final threshold satisfied max $[N_v\, N_f\, (\mu_v - \mu_f)^2]$, where N and μ were the number and mean intensity of vessel v and fundus f pixels, respectively.

The local gradient of the image, in the form of first and second directional derivatives, has also been exploited for vascular segmentation. YU *et al.* (1990), WANG and LEE (1997) and KATZ *et al.* (1988) used the first derivative, or Sobel operator, to find the magnitude and direction of the parallel edges of the vesels. They used the opposing and parallel gradients along the vessel to estimate the location of the vessel midline.

The second derivative, smoothed by a Gaussian of $\sigma = 4$, was investigated by MARTÍNEZ-PÉREZ *et al.* (1999a;b). They used the eigenvalues of the second derivatives of each pixel to obtain the maximum and minimum gradient, irrespective of their orientations. The Otsu method (OTSU, 1979) was then used to choose thresholds for the new distributions of pixel gradients after the eigenvalue transform. The means and standard deviations of the various classes obtained by these thresholds were then used to guide an iterated region growing procedure. The authors commented favourably on the applicability of this approach for images taken with the aid of contrast-enhancing agent and normal reflectance retinal photographs.

2.4.2 *Matched filters:* Matched filters are based upon similarity measures between an ideal signal and a measured signal. The sum of products of the translated matched filter over every point of the image is know as convolution and produces a probability image. A number of ideal signal profiles are available and are either based on assumptions about the anatomy and corresponding light transmittances or on empirical observation of actual vessel profiles. A number of parameters, such as length, width and orientation, describe the modelled profiles. Many other parameters are ignored; nevertheless, numerous potential matches need to be evaluated, making this approach computationally expensive in the absence of dedicated hardware.

The basic retinal vessel anatomy, represented by a circular or elliptical blood column coaxially surrounded by a thickness of wall material, has been used to generate cross-sectional profiles (RASSAM *et al.*, 1994; PATEL, 1995; HAMMER *et al.*, 1997). Other idealised profiles have been produced by the direct fitting or error minimisation of mathematical functions representing vessel profiles to the observed intensity profiles (ILIASOVA *et al.*, 1998; GAO *et al.*, 1997).

Usually, matched filters are two-dimensional cross-sections that are extruded or rotated into three dimensions. Extruded profiles are rotated to accommodate the varied orientations of the retinal vessels. Profiles that are rotated match any vessel orientation and need only to be convolved once at each image position. Matched filters that are only rotated are more susceptible to spurious detections (YANG *et al.*, 2000).

Retinal imaging varies in field of view and number of samples; parameters for the filters are thus resolution-dependent. For the extruded and rotated Gaussian filters, parameter selection has not differed from the original study (CHAUDHURI *et al.*, 1989a), with $\sigma = 2$ (GOLDBAUM *et al.*, 1989; 1993; GOH *et al.*, 1997; HOOVER *et al.*, 1998b; YANG *et al.*, 2000).

Other filter shapes have been investigated, including filters based on lines (WANG and LEE, 1997), partial Gaussians (CAN *et al.*, 1999), derivatives of Gaussians (POLI and VALLI, 1997) and ellipses (JASIOBEDSKI *et al.*, 1993). The first three of these filters consider computation time as an important design factor and rely upon binary or integer arithmetic taking less clock cycles than floating point calculations.

WANG and LEE (1997) apply filters that are 15 pixel^2, containing a line a single pixel wide, over 12 orientations. POLI and VALLI (1997) settled for a standard deviation of one, in their binomial approximation of a Gaussian. CAN *et al.* (1999) use two separate ranges of a Gaussian curve to create a matched filter corresponding to the flanks or skirts of a vascular profile.

By reducing the number of computer cycles per calculation and the total number of calculations, these studies have reduced computation time for vascular segmentation. Quantitative measures to characterise the performance of these studies are desirable, yet lacking. Matched filters do not work in isolation and form part of an algorithmic chain. They provide a non-binary measure of correlation that still needs classification into background or vessel. Thresholds divide the range of intensity of the filter response into two classes that are applied locally or globally.

HOOVER *et al.* (1998b) iteratively decrement the initial threshold (which is set to the matched filter response), so that a vessel blob is grown. This blob can be kept or discarded, depending on tests of size, relative position to already marked pixels and absolute threshold value.

Thresholds are applied to images that represent the intensity transitions of neighbouring pixels (YANG *et al.*, 2000). These image representations can be visualised as a two-dimensional array that has its number of rows and columns equal to the range of grey levels in the original image. The grey-level indices are used to accumulate the intensity transitions from each row number to column number for all neighbouring pixels in the original image.

The transformed image is useful because, when thresholded, it provides a tractable representation from which measurements can be taken to compare similarity. YANG *et al.* (2000) derived a threshold value that maximised the difference in this similarity measure and so started to segment the retinal vessels. The segmentation was finished after some of the smaller connected components (groups of pixels) were eliminated from the bilevel image.

Segmentation into vascular and non-vascular is necessary for comparison between studies or within studies to assess filter shape and parameter selection. HOOVER *et al.* (1998b) provided

the only evaluation against a human operator and reported a sensitivity (true positive fraction) of 80% and a specificity (true negative fraction) of 90% for a marked pixel matching any of its eight neighbours. The same performance was reported in a later paper (HOOVER *et al.*, 2000), presumably for a one-to-one correspondence between human-marked and computer-processed pixels.

2.4.3 *Vascular profile tracking:* From a start point on a vessel, profile tracking iteratively scans for, selects and follows pixels belonging to the retinal vascular arbour. The selection of vascular points is accomplished using two- or three-dimensional matched filters. Fuzzy vessel profile recognition and intensity based algorithms provide alternatives to matched filtering (TOLIAS and PANAS, 1998; WANG *et al.*, 1990).

The two-dimensional profile tracking algorithms reviewed used Gaussian, or a derivative of Gaussian, filters. Profiles are available that represent the vascular cross-section with less divergence in width (RASSAM *et al.*, 1994); however, these are not as analytically simple to allow least square fitting (FRAME *et al.*, 1997*a*) or for deriving indices (TAMURA *et al.*, 1988; ZHOU *et al.*, 1994) to define vessel edge and thus width.

The profiles used are only apt for describing single vessel segments, but vasculature branches and arterio-venous crossings are common. This presents a special case for tracking. FRAME *et al.* (1997*a*) used a measure of profile skewness to create new starting points for the algorithm. In a similar fashion, TAMURA *et al.* (1988) used vessel width ranges to trigger sampling of a relatively large and circular cross-sectional profile to which the filter is applied. The number of responses indicates whether branching, crossing or a vessel end point has occurred. CAN *et al.* (1999) used an asymmetry measure of the template response to trigger new tracking. KOCHNER *et al.* (1998) prospectively scanned alongside, or parallel to, the vessels that were segmented during the the first pass to detect branchings or crossings.

2.4.4 *Artificial neural network vascular segmentation:* The operation of a neural network is analogous to that of a matched filter. Both take subwindows of the image as input and return a probability measure as output. Two studies, both using the back-propagation algorithm, have detected (GARDNER *et al.*, 1996) and segmented (SINTHANAYOTHIN *et al.*, 1999) the retinal vasculature. Detection entailed classifying subwindows as containing vessels or not; segmentation classified individual pixels.

Neural networks require training examples; these examples were created by a human operator. Sinthanayothin *et al.* used 25 094 labelled (pixel-by-pixel) examples, each of 200 pixels in area. However, it was not clear whether they labelled ≈ 2.5 million pixels to create independent and non-overlapping examples, or whether they labelled an area that could generate 25 094 examples whose input (but not output) areas overlapped.

The neural networks researched by GARDNER *et al.* (1996) used 20×20 pixel subwindows. Nine thousand of these subwindows were marked for neural learning validation. Generalisation assessment over 1200 unseen subwindows resulted in a sensitivity and specificity of 91.7%. SINTHANAYOTHIN *et al.* (1999) achieved a sensitivity of 83% and a specificity of 91% in classifying pixels.

3 Image processing for diabetic retinal pathology

The grading methods of the early treatment diabetic retinopathy study (ETDRS) described the assessment of at least 22 categories of lesions with respect to DR (ETDRS RESEARCH GROUP, 1991*a*). There are other, less formal, grading concepts

used in DR assessment, adding to the number of lesion categories. Grading concepts, such as ischaemia and 'dots and blots', have not been assessed in the early treatment trials for prognostic significance.

When this variation is combined with the natural variability of the image and the variety of confounding non-diabetic pathologies, it makes segmenting lesions a difficult task (DONOHOE *et al.*, 1999). Research has focused on the identification of exudates, micro-aneurysms and haemorrhages, which are relatively numerous and easier to define in terms of area and perimeter.

3.1 *Registration of retinal images*

If screening incorporates digital imaging of the retina, multiple images are usually recorded for each eye at examinations. The magnification and location of each image differ. In systematic screening, the locations of these images on the retina, as well as the field of view, are defined by the screening protocol, and variability in these protocols is compensated for by the human interpreter.

Digital analysis, in an analogous fashion to human interpretation, seeks to standardise the retinal areas of interpretation. Registration, or finding a single spatial and temporal co-ordinate set, enables the repeatable analysis of exactly the same scene or domain in different fundus images of the same eye. Registration can thus monitor changes and differences in a given area of retina and also enables integration of images to cancel noise or to create wide-angle image montages.

The unknown geometric distortion arising from the lens and the spherical nature of the retina introduce some non-linear distortions. However, the assumption of a planar area of interest, combined with relaxed requirements for the fidelity of the geometrical transform, enables rotation, scaling and translation (RST) distortions to be modelled.

Many methods have been proposed for registration, but a demonstration of their potential for automated registration outside laboratory conditions remains (WOOD, 1990; GOH *et al.*, 1997; RYAN and HENEGHAN, 1999). The difference between precisely registered serial images was used as an indicator of developed non-diabetic pathology by CIDECIYAN *et al.* (1992) and GOLDBAUM *et al.* (1993). However, no studies have been found expanding upon this methodology.

BALLERINI (1998) presents a method to compensate for motion and the changes in contrast resulting from the fluorescent dye travelling down the retinal and subretinal vasculature. This could have applicability in imaging modalities for screening, where annular or crescent-shaped areas of low contrast move from image to image.

Other registration tasks have been developed for motion compensation or tracking. DOMINGO *et al.* (1997) investigated motion recovery in variable-contrast angiograms of the retina. Albeit not directly applicable to screening, investigations managing and incorporating variable and 'badly behaved' marker points into algorithms for the estimation of the geometrical transform are useful. They are useful because a significant proportion of each retinal image is irregular, owing to previous therapy, non-diabetic pathology or to factors (described in Section 2.2) intrinsic to the imaging process. Real-time systems used to locate, guide and verify the placement of the laser for photocoagulation of the retina may one day be incorporated into image acquisition systems for screening. Details of the algorithms and performance characteristics are given in Table 2.

3.2 *Micro-aneurysm segmentation*

Micro-aneurysms are spherical balloonings of the vascular wall; their segmentation is based upon elimination of linear

Table 2   Overview of registration approaches

Study	Transform	Error	Markers	Summary of algorithm
GOLDBAUM *et al.* (1993)	RST	registration failed in 16%	matched filter response	correlation of POI to form feature pairs; iterative least squares on decreasing best-fit percentiles
SPENCER *et al.* (1996)	non-linear	NA	human-marked anatomy	VisiLog
GOH *et al.* (1997)	translation	nine image montage accrued errors of 5 pixels	whole image	error-threshold-restricted global search by translation of images
RYAN and HENEGHAN (1999)	RST	sum squared of intensity difference was 1%	human-marked anatomy	POI pairs form vectors; metrics describing tentative R, S and T transforms for all vector pairs between images are plotted; largest cluster is taken as correct transform
DOMINGO *et al.* (1997)	RST	MSE 0.84–16.6 pixels. Failed in five out of 52 owing to POI selector	POIs selected by Förstner's operator	approximation of correspondence pairs based on Euclidian distance between feature vectors; followed by either iterative refinement or a least squares error by discarding points furthest from match
TANAKA and TANAKA (1988)	translation	NA	vessel branching points	correlation between features of branching points (length and direction of branches); matching points are plotted on a scatter diagram; largest clusters are taken to represent translation
BALLERINI (1998)	translation	NA	matched filter response	cross-correlation of binary images.
BECKER *et al.* (1998)	translation and scaling	average 1.35 pixels, always below 4 pixels	points of edge direction dispersion	disperion measure of edge strength-direction histogram is used to reject candidate points; best five matches start clustering for remaining points; candidate transform is validated using an empirically set normalised sum-of-differences between images and minimised via least squares
BARRETT *et al.* (1994)	translation	NA	vessel cross-section templates	small templates are scanned over a constrained area derived from maximum theoretical displacements
CIDECIYAN *et al.* (1992)	RST	average pixel intensity SD = 4.65	whole image	log-polar representation of image, in Fourier domain, allows scaling and rotation to be represented as shifts that are amonable to a 'brute force' cross-correlation approach

MSE = mean squared error; NA = none available; POI = points of interest; RST = rotation scale translation

structures and subsequent feature-based discrimination of candidate areas. Exudates, which are lipid-like deposits, are segmented by thresholding in alternate colour spaces. Texture information is used to classify larger regions of pathology, such as new vessel growth and ischaemia. The fractal dimension, representing the degree to which the structure fills the dimensions in which it is embedded, is also used to study new vessel growth.

Micro-aneurysms range in size between $12\mu m$ and $125\mu m$. The small size of micro-aneurysms means that image processing is usually applied to images obtained with the aid of a blood-borne contrast agent. However, these methods and techniques have also been applied, with some success, to conventional reflectance images.

Micro-aneurysm segmentation usually started with the subtraction of an image containing linear features. These features were detected by erosion and subsequent dilation with a linear structuring element (PHILLIPS *et al.*, 1991; SPENCER *et al.*, 1994; 1996; FRAME *et al.*, 1996; CREE *et al.*, 1997; MENDONÇA *et al.*, 1999). The resulting image was match filtered (using an 11×11 pixel 2D Gaussian filter of $\sigma = 1$) and thresholded to obtain seed points. Pixels were added to the seed-points to form regions, the inclusion criteria being based upon an empirical contrast ratio between the background and the intensity of the seed pixel. Parameters such as area, perimeter and energy (sum of squared intensities) were used to described each region.

Humanly delineated micro-aneurysms enabled decision boundaries to be drawn on distribution plots of these parameters; segregating the candidate regions into presumed micro-aneurysm and spurious artifact. SPENCER *et al.* (1996) achieved a sensitivity of 82% at a cost of detecting between 100 and 150 false positives over four images. CREE *et al.* (1997) improved on this by using more parameters and reduced the false positive rate to 5.7 per image for the same sensitivity. A sensitivity of 66% was achieved by EGE *et al.* (1998), with 107 false positives from a sample containing 316 micro-aneurysms.

3.3 *Exudate and haemorrhage segmentation*

Exudates, like micro-aneurysms, are an early indicator of diabetic retinal disease. Human selection of the yellow and waxy sheened pixels representative of exudate areas was used to plot scattergrams of colour features of these pixels (GOLDBAUM *et al.*, 1990*b*). More specifically; a brightness-independent representation of exudate colour was plotted with the same representation of haemorrhage and infarct lesions. Boundaries set by human operators distinguished the three types of lesion on these

scattergrams. The system was only tested to classify pixels that had already been selected by a human. Exudates were thus distinguished from other focal lesions in seven out of ten cases.

WARD *et al.* (1988) presented an empirical and interactive global technique for thresholding a noise-reduced (mean-filtered, 64 pixel^2) image to segment exudates. A fully auto-mated method applied a 'simple threshold' after the vessels and OD had been removed (KOCHNER *et al.*, 1997). 'Simple thresh-olding' was also used by ZAHLMANN *et al.* (2000) and was based upon an initial threshold value that segmented 5% of the brightest pixels. This was used to probe, under empirical area constraints, for more exudate pixels. The edge responses of the potential exudative pixels and their surrounding neighbourhood provided further basis for discrimination.

EGE *et al.* (1998) detected all but ten of 390 exudative lesions to give a sensitivity of 97%; no details were given of the algorithm. In a later paper (EGE *et al.*, 2000), a region-growing approach was used upon seed points derived from a background-subtracted image. A median filter, 31 pixel^2, allowed thresholds to be set to segment partially bright lesions such as exudates. An empirically constrained region-growing algorithm then further segmented the candidate areas. Features extracted from this region were assessed for significance and presented to three classification methods. A nearest neighbour-based classification was able to discern all exudates (100% sensitivity) in 134 images.

HUNTER *et al.* (2000) evaluated abstract features derived from statistics and transforms summarising subwindows of the image. From a data set of 16 images, 211 subwindows, each 16 pixel^2, were found and marked as containing exudates or drusen (which are similar to exudates, but are more recessed and punctuate in appearance). For each of these subwindows, 178 features were developed and were based on the raw image, the Fourier transform and the directional derivatives. From these 178 features, it was found that an input feature vector of 11 data points was necessary to achieve a sensitivity and specificity of 91% in the detection of these lesions.

Texture-based segmentation was applied by FRAME *et al.* (1997b): exudates were marked by a human, and their intensity histogram and their spatially dependent intensity transition (an indicator of texture) probabilities were computed. A Gaussian form was assumed for the distribution of metrics describing the exudates and normal background. Prospective exudate pixels were classified according to the distance from the mean of the humanly marked ones. Classification performance was described as inferior to that of interactive global thresholding.

Haemorrhages have irregular boundaries and are either flame-or feather-like in texture and shape. They range in size from relatively extensive (occupying quartiles of an image) to small and approaching the size of micro-aneurysms. LEE *et al.* (1999) used sequential or staggered global thresholding and a sampling grid formed from the intersection of radial lines and concentric circles converging to the macula for segmentation. They described their results as 'manually verified and found in good agreement with the manual counts'.

Haemorrhage detection by EGE *et al.* (1998) was successful in 140 out of 155 lesions. Neural networks have also been applied to the detection of haemorrhages. More than 12 000 subwindows (30 pixel^2) were used to train and test a back-propagation network that achieved a sensitivity and specificity generalisation ability of 73.8% (GARDNER *et al.*, 1996).

BERRY and WESTERMAN (1998) proposed a clustering algo-rithm for the identification of micro-aneurysms, haemorrhages and the vasculature. The assignment of a lesion type to the clusters was based on the major and minor axes of the bounding ellipse and an intensity-modulated measure of spread, termed the radius of gyration. These metrics seemed to vary with lesion type but were not integrated or applied to the segmentation of haemorrhages. A similar approach was proposed by LEE and

WANG (1998), where a range of thresholds provided binary images for various matched filters corresponding either to vessels or punctuate or diffuse lesions.

3.4 *Analysis of the ischaemic and neovascular retina*

The capillaries of the retina help determine the texture and appearance of the fundus. The presence of new vessels is associated with a high risk of progression to the sight-threa-tening stages of diabetic retinopathy. Likewise, atrophy of the normal capillary bed, causing ischaemia, indicates progression at an earlier stage of pathology. The border of the capillary bed surrounding the area of greatest visual acuity is mildly curved in shape; disturbances in this boundary region indicate changes in the distribution of the capillary bed.

JASIOBEDSKI *et al.* (1993), working with contrast-enhanced retinal images, compared an ischaemic classification algorithm with classifications by an ophthalmologist. Their segmentation approach was singular in that pixels were amalgamated into homogenous regions. These were created by smoothing, erosion followed by dilation and a reduction in the number of grey levels. From the resulting patchwork or tessellation of regions and their borders, the vessels were detected by matched filtering (with a semi-elliptical profile) normal to the borders. This enabled vessel pixels to be excluded from metrics describing the region's interior texture. The change in intensity in the patches that occurred with different structuring elements used in the erosion and dilation described the abundance and extent of features within that region. The system reached a specificity and sensitivity of approximately 80% compared with the classifica-tions given by an ophthalmologist.

Ischaemia can lead to new vessel growth and changes to the avascular zone of the fovea. FRAME *et al.* (1997b) used the intensity histogram to derive texture metrics for neovascular areas. Regions with and without neovascularisation were selected and marked by humans and provided a training set to optimise the weights for the classifier (linear discriminant analysis). The human-marked classes were linearly separable, but applicability to new data was found to be disappointing.

The fractal dimension of a retinal region gives a scale-invariant description of how features fill that region. Measures of the fractal dimension have been applied to study patterns of neovascularisation. Although not true fractals (PANICO and STERLING, 1995; WEST *et al.*, 1997) and better described as space filling, vessel segment lengths are clearly disturbed during neovascularsation.

The fractal dimension, as measured in a region of interest (ROI) using the density–density correlation method, has been shown to change for a patient over the course of neovascularisa-tion (DAXER, 1993a) and between normal and neovascularised retinae (DAXER, 1993b). LANDINI *et al.* (1993) measured the fractal dimension for whole retinal images of normal retinas and for those with venous oclusion (LANDINI *et al.*, 1995) and found no difference. However, the fractal dimension of connected components within subwindows distinguishes pathological and normal states. Unfortunately, the fractal studies segmented the retinal vasculature by manually tracing the vessel midlines on paper and so lost their applicability for automation.

The rounded transition zone from vascular to avascular retina surrounding the foveola has been delineated with contours fitted using genetic algorithms. Parameters describing the flexibility and length of the contour were tweaked (BALLERINI, 1999). The position and shape of the contour were then found within a pre-centred annular area on the fovea by mutating, recombining and selecting the best solutions over many generations of trials. Two sample contours were presented.

4 Intelligent systems for DR analysis, diagnosis and screening

A modular intelligent system for DR diagnosis would identify and segment each type of lesion; so that, by reference to a grading scheme, the level of DR could be ascertained. In practice, once some pathology has been segmented, attempts are made to use that information. For example, the absence or questionable presence of micro-aneurysms would not require immediate intervention. Exudates and haemorrhages have also been used to provide information for diabetic ocular health care.

GARDNER et al. (1996) used a neural network to classify subwindows as vessel, haemorrhage or exudate. With this information, the system was compared with the referral rates of an ophthalmologist. The Gardner et al. system would not have missed any patient that required referral (99% sensitive) and was 69% specific; in other words, 31% of all diabetics not requiring attention would have been flagged for referral. No details were given on how the information on exudates and haemorrhages was encoded and integrated to come to this decision.

ALEYNIKOV and MICHELI-TZANAKUO (1998) described a neural system with input features based on the spatial frequencies present within a sub-image. They also used various averages of these subwindows to form the inputs to a neural network that had targets derived from a human classification of the severity of haemorrhages. System performance was poorly specified; the reported accuracy of 79% does not give information on the classification of true negatives and true positives for a given prevalence.

HIPWELL et al. (2000) evaluated an intelligent system for screening DR. A human grader classed 3783 images with reference to standard photographs, according to the EURDIAB protocol (ALDINGTON et al., 1995). Three aspects of this grading scale were compared with the micro-aneurysm detector (SPENCER et al., 1996; CREE et al., 1997) described in Section 3.2. The micro-aneurysm segmentation algorithm, operating as an intelligent screening tool, was set to give a positive result if any micro-aneurysm was detected. A comparison with the human grader is summarised in Table 3. This system could reject 76% of all patients with no DR at the expense of rejecting 15% of patients who actually have DR (according to the grading methodology and criteria used).

HIPWELL et al. (2000) demonstrated how, for a given grading methodology, screening based upon segmentation of a single feature can save diabetic health care resources. ZAHLMANN et al. (2000) proposed an elaborate system that encompassed some of the non-imaging aspects of DR grading and screening. Inputs describing the patients' metabolic profile, treatment history and visual acuity were combined with parameters from retinal image segmentation. The vascular arbour (segmented as in KOCHNER et al. (1998)) was obtained and was used as a framework from which to identify the OD and foveal area, but it was not used per se in subsequent intelligent analyses. The exudative area, described in Section 3.3, and all the other inputs were given a degree of significance by a medical expert. Fuzzy rules, tuned using simulated data, were tested upon 12 real patient data sets, eight of which had retinopathy not detectable using non-stereoscopic retinography alone i.e. macular oedema. The system classified correctly 96% of image-based decisions and 76% of cases partially involving non-photographic data. No break-down was given of the type of truth (true positives or true negatives) for which the system was correct.

There have been many other abstracts and poster presentations on automated screening for DR. Implementation details and performance results are less detailed than those reviewed in this Section. Table 4 summarises their approaches and claims.

Table 3 *Prospective validation of micro-aneurysm detection for DR screening with 3783 images or 925 studies. 174 images were classified by human grader into questionable category for micro-aneurysm present; these were not comparable with output from micro-aneurysm detection algorithm. All data are from* HIPWELL et al. (2000)

	Single image		Complete study
	Micro-aneurysms (number of cases detected)	DR (number of cases detected)	DR (number of cases detected)
Sensitivity, %	81 (736 913)	78 (745 956)	85 (315 372)
Specificity, %	93 (2 502 2696)	91 (2 576 2827)	76 (423 553)

Table 4 *Communications presenting intelligent systems for DR analysis*

Study	Methods	Presented outcome
ZAHLMANN et al. (1996)	no details	proposal of interactive, on-line DR atlas and diagnostic query tool
NGUYEN et al. (1996)	neural classification of segmented lesions	human-labelled images compressed and automatically graded according to ETDRS RESEARCH GROUP (1991a) using back-propagation neural networks
CASI et al. (1996)	fuzzy neural	graphical user interface to ease access to HIS, PACS and intelligent diagnostic modules
SINCLAIR et al. (1996)	no details	automated classification (conforming to early version of ETDRS RESEARCH GROUP (1991a)) claimed to compare well with that of retinal specialist; third grading standard is implied but not specified
GOLDBAUM et al. (1996)	no details	database of images is presented that can be queried by lesion type and thus retinopathy grade; querying mechanism works upon automatically generated index
THOMPSON (1999)	neural analysis of wavelet-compressed images	images from 150 subjects (256×256 pixels) reduced to 176 orthogonal wavelet coefficients; neural classification of coefficients was 83% specific and sensitive
DONOHOE et al. (1998)	neural analysis with interactive segmentation of lesions	neural labelling of lesions guided by human operator enables semi-quantitative labelling of fundus areas

5 Summary and discussion

Owing to the scarcity of objective results, only a few of the studies reviewed are worthy of note. For anatomical image-analysis, neural- and variance-based algorithms for the segmentation of the vasculature and OD have effective performance (SINTHANAYOTHIN *et al.*, 1999). The vascular segmentation results of HOOVER *et al.* (2000) are remarkable for the use of a representative data set and may be more relevant and reliable with respect to screening situations. In these non-ideal environments, image quality, previous therapy and non-diabetic lesions will interfere with analysis unless these images are pre-emptively flagged, as suggested by LEE and WANG (1999).

The intelligent system proposed by HIPWELL *et al.* (2000) demonstrated that diabetic patients can be reliably dichotomised into those with definitely no micro-aneurysms and those with possible micro-aneurysms. This was achieved by empirical means, whereby the properties of human-labelled micro-aneurysms could be plotted with candidate areas produced by the algorithm. In this feature, space decision boundaries could be drawn. Screening on the basis of establishing the absence of micro-aneurysms is an important first step, but it does not reliably identify the diabetic patient with no DR.

The image types, patient information and lesions required for medical management decisions for DR are well-defined, clearly demonstrated and well-disseminated. It is clear that digital imaging can only record a subset of this data. To compare results of lesion or anatomy segmentation for intelligent decisions or otherwise, a defined standard needs to be referenced. Pre-selection of the test images, local standards for DR grading and qualitative description of results do not allow objective comparisons but are common in the literature (GOLDBAUM *et al.*, 1989; CASI *et al.*, 1996; NGUYEN *et al.*, 1996; ZAHLMANN *et al.*, 1996; 2000; ALEYNIKOV and MICHELI-TZANAKUO, 1998; BERRY and WESTERMAN, 1998).

The gold-standard method for grading and describing DR is seven-field stereoscopic photography. This is a descriptive and labour-intensive grading protocol that enables effective timing and early treatment to prevent or delay sight loss (DRS RESEARCH GROUP, 1987; ETDRS RESEARCH GROUP, 1985; 1991*b*).

Novel grading schemes are validated against this grading method (KLEIN *et al.*, 1986; ALDINGTON *et al.*, 1995), and epidemiological studies use the detail it provides to monitor incidence and the effects of therapy (KLEIN *et al.*, 1989; DCCT RESEARCH GROUP, 1995). There are thus a well-established diagnostic framework and objective criteria by which pathology segmentation algorithms or intelligent systems can be evaluated.

However, there is still debate as to whether digital imaging is capable of capturing enough, or all, of the required information. The aspect of DR know as macular oedema can only be detected if stereoscopic images are taken or by direct examination, such as slit-lamp biomicroscopy. The detection of drusen is also aided by stereopsis. AIELLO *et al.* (1998*b*) took stereopsis into account when assessing digital imaging for DR diagnosis. The study found that 30° stereoscopic grading of DR using digital images according to the ETDRS protocol was feasible.

Non-stereoscopic digital-imaging screening has also been favourably proposed (YOUNG *et al.*, 1997; GEORGE *et al.*, 1997; LIESENFELD *et al.*, 2000; HENRICSSON *et al.*, 2000). The lower resolution of digital images compared with that of photographic film was not found to hamper stereoscopic DR diagnosis and the detection of drusen (AIELLO *et al.*, 1998*b*; SOLIZ *et al.*, 2000). Topographic images digitally representing the baseline disparity have been proposed by BERESTOV (2000) and hold great potential for the analysis of macular oedema.

The inclusion of macular oedema into automated screening is important and relevant, because a significant proportion, ≈10%

(KLEIN *et al.*, 1989), of sight-threatening cases would be missed by the segmentation techniques and developing intelligent screening systems reviewed in this paper, even at a sensitivity of 100%. It is precisely for patients with macular oedema that early treatment is indicated, because it is for these patients that sight loss can be prevented as opposed to halted (ETDRS RESEARCH GROUP, 1985).

Image analysis is starting to address the constraints and implications found in the underlying data generation mechanisms found in actual diabetic screening. HOOVER *et al.* (2000) and LEE and WANG (1999) have integrated problems that would badly affect most of the algorithms presented. Issues such as non-diabetic lesions, previous treatment and different imaging protocols are all still confounding factors. Image quality has also been identified as a limiting factor (HIPWELL *et al.*, 2000; EGE *et al.*, 2000).

The few research papers investigating full automation for screening emphasise that detection of pathology should trigger further referral; high sensitivities are required to ensure false negatives do not occur, and this approach lightens the burden on screening personnel by flagging abnormal images. Inverting the screening task, i.e. detecting normal diabetic retinae, would accomplish the same result, releasing safe patients back into the screening pool and flagging suspect cases of all pathology types.

The concerns regarding the subjective reporting of results are aptly put forward by PELI (1993), who highlights the need for the objective evaluation of the efficacy of systems. Although Peli discusses image enhancement, the lack of validity in transferring evidence of improved performance in the laboratory to a clinical situation needs careful consideration within automated screening as well, as it has been found that computer systems aiding in the diagnosis are not as effective as simple and objective interpretative aids, such as formal protocols and reference diagrams. This improvement has also been borne out in studies for the automation of diagnosis of pulmonary embolisms (FISHER *et al.*, 1996; SCOTT *et al.*, 1996).

Further parallels between screening for pulmonary embolisms, cardiac infarcts and DR are notable. Many intermediary representations or features that have no analogue in modular, or lesion-by-lesion, grading can be exploited for DR analysis. BAXT (1994), BAXT and SKORA (1996) and the aforementioned pulmonary embolism studies echo the work of HUNTER *et al.* (2000) and ALEYNIKOV and MICHELI-TZANAKUO (1998) on the use of unconventional prognosticators of disease. However, the analysis of the significance of single features by HUNTER *et al.* (2000) is likely to ignore the non-linear interactions so fruitfully exploited by BAXT and SKORA (1996).

One of the more promising indices of DR is described by KRISTINSSON *et al.* (1997), who have uncovered an interesting correlation between vascular segment length and developing macular oedema. The algorithms presented by GREGSON *et al.* (1995); KOCHNER *et al.* (1998); HOOVER *et al.* (2000) and SINTHANAYOTHIN *et al.* (1999) are well positioned, with a little modification, to extract segment length and calibre information and so include the diabetic population for which the timing of the detection and treatment of asymptomatic DR lesions is so critical.

References

AIELLO, L., CAVALLERANO, J., GARDNER, T., KING, G., BLANKENSHIP, G., FERRIS, F., and KLEIN, R. (1998*a*): 'Diabetic retinopathy', *Diabetes Care*, **21**, pp. 143–156

AIELLO, L., BURSELL, S., CAVALLERANO, J., KELLY GARDNER, W., and STRONG, J. (1998*b*): 'Joslin vision network validation study: Pilot image stabilization phase', *J. Am. Optometr. Assoc.*, **69**, pp. 699–710

AKITA, K., and KUGA, H. (1979): 'Towards understanding color ocular fundus images'. *Proc. 6th IJCAI*, Tokyo, Japan, pp. 7–12

AKITA, K., and KUGA, H. (1982): 'A computer method of understanding ocular fundus images', *Pattern Recognit.*, **16**, pp. 431–443

ALDINGTON, S., KOHNER, E., MEUER, S., KLEIN, R., and SJOLE, A. (1995): 'Methodology for retinal photography and assessment of diabetic retinopathy: the EURODIAB IDDM complications study', *Diabetologia*, **38**, pp. 437–444

ALEYNIKOV, S., and MICHELI-TZANAKUO, E. (1998): 'Classfication of retinal damage by a neural network based system', *J. Med. Syst.*, **22**, pp. 129–136

AMERICAN DIABETES ASSOCIATION (1999): 'American diabetes association: Clinical practice recomendations 1999'. Technical Report S1, American Diabetes Association

BALLERINI, L. (1998): 'Integration of retinal image sequences'. Proceedings of SPIE Conference on Applications of Digital Image Processing XXI, **3460**, pp. 237–248

BALLERINI, L. (1999): 'Detection and quantification of diabetic retinopathy'. Proceedings of SPIE Conference on Applications of Digital Image Processing XXII, **3808**, pp. 213–223

BARRETT, S., JERATH, M., RYLANDER III, H., and WELCH, A. (1994): 'Digital tracking and control of retinal images', *Optical Eng.*, **33**, pp. 150–159

BAXT, W. G. (1994): 'A neural network trained to identify the presence of myocardial infarction bases some decisions on clinical associations that differ from accepted clinical teaching', *Med. Decis. Making*, **14**, pp. 217–222

BAXT, W. G., and SKORA, J. (1996): 'Prospective validation of artificial neural network trained to identify acute myocardial infarction [see comments]', *Lancet*, **347**, pp. 12–15

BECKER, D., CAN, A., TURNER, J., TANNENBAUM, H., and ROYSAM, B. (1998): 'Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination', *IEEE Trans. Biomedical Eng.*, **45**, pp. 105–117

BERESTOV, A. (2000): 'Stereo fundus photography: automatic evaluation of retinal topography'. Proceedings of SPIE Conference on Stereoscopic Displays and Virtual Reality Systems VII', pp. 50–59

BERRY, M., and WESTERMAN, D. (1998): 'Cluster form analysis techniques for diabetic retinopathy', *Proc. Math. Models Med. Health Sci.*, pp. 35–50

CAN, A., SHEN, H., TURNER, J., TANENBAUM, H., and ROYSAM, B. (1999): 'Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms', *IEEE Trans. Inform. Technol. Biomed.*, **3**, pp. 125–138

CASI, E., CERAVOLA, A., CIONINI, R., SPERDUTI, A., STARITA, A., and VITI, S. (1996): 'Diabetic retina analyser' in BOOM, H. (Ed).: 'Proceedings of 18th Annual International Conference of IEEE Engineering in Medicine and Biology Society' (Amsterdam, The Netherlands), pp. 1128–1129

CHAUDHURI, S., CHATTERJEE, S., KATZ, N., NELSON, M., and GOLDBAUM, M. (1989*a*): 'Detection of blood vessels in retinal images using two-dimensional matched filters', *IEEE Trans. Med. Imag.*, **8**, pp. 263–269

CHAUDHURI, S., CHATTERJEE, S., KATZ, N., and GOLDBAUM, M. (1989*b*): 'Automatic detection of the optic nerve in retinal images'. IEEE International Conference on Image Processing, Singapore, Vol. 1, pp. 1–5

CIDECIYAN, A., JACOBSON, S., KEMP, C., KNIGHTON, R., and NAGEL, J. (1992): 'Registration of high resolution images of the retina', Proceedings of SPIE Conference on Medical Imaging VI: Image Processing, **1652**, pp. 310–322

CREE, M., OLSON, J., MCHARDY, K., SHARP, P., and FORRESTER, J. (1997): 'A fully automated comparitive microaneurism digital detection system', *Eye*, **11**, pp. 622–628

DASBACH, E., FRYBACK, D., NEWCOMB, P., KLEIN, R., and KLEIN, B. (1991): 'Cost-effectiveness of strategies for detecting diabetic retinopathy', *Med. Care*, **29**, pp. 20–39

DAXER, A. (1993*a*): 'Characterization of the neovascularisation process in diabetic retinopathy by means of fractal geometry: diag-

nostic implications', *Graefe's Arch. Clin. Exp. Ophthalmol.*, **231**, pp. 681–686

DAXER, A. (1993*b*): 'The fractal geometry of proliferative diabetic retinopathy: implications for the diagnosis and the process of retinal vasculogenesis', *Current Eye Res.*, **12**, pp. 1103–1109

DCCT RESEARCH GROUP (1995): 'Progression of retinopathy with intensive versus conventional treatment in the Diabetes Control and Complications Trial', *Ophthalmology*, **102**, pp. 647–661

DIABETES UK (2000): '*Diabetes in the UK—the missing million*'. The British Diabetes Association

DOMINGO, J., AYALA, G., SIMO, A., DEVES, E., MARTINEZ COSTA, L., and MARCO, P. (1997): 'Irregular motion recovery in fluorescein angiograms', *Pattern Recognit. Lett.*, **18**, pp. 805–821

DONOHOE, G., SOLIZ, P., and NEMETH, S. (1998): 'Computer-aided image analysis for background diabetic retinopathy'. Proceedings of SPIE Conference on Medical Imaging 1998: Image Processing, **3338**, pp. 1017–1027

DONOHOE, G., NEMETH, S., and SOLIZ, P. (1999): 'An interactive system for computer-aided retinal image analysis'. 12th IEEE Symposium on Computer Based Medical Systems, 306, Los Alamitos, pp. 184–189

DRS RESEARCH GROUP (1981): 'Report 7. a modification of the airlie house classification of diabetic retinopathy', *Invest. Opthalmol. Vis. Sci.*, **21**, pp. 210–226

DRS RESEARCH GROUP (1987): 'Indications for photocoagulation treatment of diabetic retinopathy: Diabetic retinopathy study report number 14', **27**, pp. 239–253

EGE, B., HEJLESEN, O., LARSEN, O., JENNINGS, B., KERR, D., and CAVAN, D. (1998): 'Screening for diabetic retinopathy using computer based image analysis and bayesian classification', *Diabetes, Nutrit. Metabol.*, **11**, p. 95

EGE, B., HEJLESEN, O., LARSEN, O. V, M, JENNINGS, B., KERR, D., and CAVAN, D. (2000): 'Screening for diabetic retinopathy using computer based image analysis and statistical classification', *Comput. Methods Progr. Biomed.*, **62**, pp. 165–175

ETDRS RESEARCH GROUP (1985): 'Photocoagulation for diabetic macular edema: Early treatment diabetic retinopathy study report number 1', *Arch. Ophthalmol.*, **103**, pp. 1796–1806

ETDRS RESEARCH GROUP (1991*a*): 'Grading diabetic retinopathy from stereoscopic color fundus photographs- an extension of the modified Airlie House classification: Early treatment diabetic retinopathy study report number 10', *Ophthalmology*, **98**, pp. 786–806

ETDRS RESEARCH GROUP (1991*b*): 'Early photocoagulation for diabetic retinopathy: Early treatment diabetic retinopathy study report number 9', *Ophthalmology*, **98**, pp. 766–785

FISHER, R., SCOTT, J., and PALMER, E. (1996): 'Neural networks in ventilation-perfusion imaging; part 1. Effects of interpretive criteria and network architecture', *Radiology*, **198**, pp. 699–706

FRAME, A., CREE, M., OLSON, J., UNDRILL, P., MCHARDY, K., SHARP, P., and FORRESTER, J. (1996): 'Computer based classification of retinal micro-aneurysms'. International Conference on Neural Networks and Expert Systems in Medicine and Health Care, Plymouth, United Kingdom, pp. 50–56

FRAME, A., UNDRILL, P., OLSON, J., MCHARDY, K., SHARP, P., and FORRESTER, J. (1997*a*): 'Structural analysis of retinal vessels'. Proceedings of 6th International Conference on Image Processing and its Applications, IEE, Dublin, Vol. 2, pp. 824–827

FRAME, A., UNDRILL, P., OLSON, J., MCHARDY, K., SHARP, P., and FORRESTER, J. (1997*b*): 'Texture analysis of retinal neovascularisation'. Proceedings of IEE Colloquium on Pattern Recognition, London, Vol. 6, pp. 1–6

GAO, X., BHARATH, A. A., HUGHES, A., A., S., CHAPMAN, N., and THOM, S. (1997): 'Towards retinal vessel parameterisation'. Proceedings of SPIE Conference on Medical Imaging, **3034**, pp. 734–744

GARDNER, G., KEATING, D., WILLIAMSON, T., and ELLIOT, A. (1996): 'Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool', *Br. J. Ophthalmol.*, **80**, pp. 940–944

GEORGE, L., LEVERTON, C., YOUNG, S., LUSTY, J., DUNSTAN, F., and OWENS, D. (1997): 'Can digitised colour 35 mm transparencies be used to diagnose diabetic retinopathy?', *Diabet. Med.*, **14**, pp. 970–973

GOH, K., SARKODIE-GYAN, T., CAMPBELL, A., SIMPSON, D., and MCNEELA, B. (1997): 'Computer assisted photocoagulation for treatment of diabetic retinopathy:preliminary results'. Proceedings of 30th International Symposium on Automotive Technology and Automation, Florence, Vol. 2, pp. 655–662

GOLDBAUM, M., KATZ, N., CHAUDHURI, S., and NELSON, M. (1989): 'Image understanding for automated retinal diagnosis'. Proceedings of IEEE Symposium for Computer Applications in Clinical Medicine, pp. 756–760

GOLDBAUM, M., KATZ, N., CHAUDHURI, M., NELSON, M., and KUBE, P. (1990a): 'Digital image processing for ocular fundus images', *Ophthalmol. Clin. North Am.*, **3**, pp. 447–466

GOLDBAUM, M., KATZ, N., NELSON, M., and HAFF, L. (1990b): 'The discrimination of similarly colored objects in computer images of the ocular fundus', *Invest. Ophthal. Vis. Sci.*, **31**, pp. 617–623

GOLDBAUM, M., KOUZNETSOVA, V., COTE, B., HART, W., and NELSON, M. (1993): 'Automated registration of digital ocular fundus images for comparison of lesions', *SPIE: Ophthalmic Technol. III*, **1877**, pp. 94–99

GOLDBAUM, M., JAIN, R., GUPTA, A., MOEZZI, S., TAYLOR, A., CHATTERJEE, S., and BURGESS, S. (1996): 'Database for ocular fundus images ranked by semantic similarity', *Investigat. Ophthalmol. Vis. Sci.*, **37**, p. S957

GONZALEZ, R., and WOODS, R. (1992): 'Digital image processing' (Addison-Wesley, 1992)

GREGSON, P., SHEN, Z., SCOTT, R., and KOZOUDEK, V. (1995): 'Automated grading of venous beading', *Comput. Biomed. Res.*, **28**, pp. 291–304

HAMMER, M., LEISTRITZ, S., LEISTRITZ, L., SCHWEITSER, D., THAMM, E., and DONNERHACKE, K. H. (1997): 'Monte-Carlo simulation of retinal vessel profiles for the interpretation of *in vivo* oxymetric measurements by imaging fundus reflectometry'. Proceedings on Medical Applications of Lasers in Dermatology, Ophthalmology, Dentistry, and Endoscopy, San Remo, Italy, Vol. 3192, pp. 211–218

HART, W., GOLDBAUM, M., COTE, B., KUBE, P., and NELSON, M. (1997): 'Automated measurement of retinal vascular tortuosity'. Proceedings of American Medical Informatics Association Annual Fall Symposium, Nashville, Vol. 63, pp. 459–463

HENRICSSON, M., KARLSON, C., EKHOLM, L., KAIKKONEN, P., SELLMAN, A., STEFFERT, E., and TYRBERG, M. (2000): 'Colour slides or digital photography in diabetes screening—a comparison', *Acta Ophthalmologica Scand.*, **78**, pp. 164–168

HIPWELL, J., STRACHNT, F., OLSON, J., MCHARDY, K., SHARP, P., and FORRESTER, J. (2000): 'Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool', *Diabet. Med.*, **17**, pp. 588–594

HOOVER, A., and GOLDBAUM, M. (1998a): 'Fuzzy convergence'. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, pp. 716–721

HOOVER, A., KOUZNETSOVA, V., and GOLDBAUM, M. (1998b): 'Location blood vessels in retinal images by piece-wise threshold probing of a matched filter response', American Medical Informatics Association Conference, Orlando, pp. 931–935

IIOOVER, A., KOUZNETZOVA, V., and GOLDBAUM, M. (2000): 'Locating blood vessels in retinal images by piece-wise threshold probing of a matched fileter response', *IEEE Trans. Med. Imaging.*, **19**, pp. 203–210

HUNTER, A., LOWELL, J., OWENS, J., KENNEDY, L., and STEELE, D. (2000): 'Quantification of diabetic retinopathy using neural networks and sensitivity analysis'. Proceedings of Artificial Neural Networks in Medicine and Biology, Götenborg, Sweden, pp. 81–86

ILIASOVA, N., USTINOV, A., BRANCHEVSKY, S., and DURASOV, A. (1998): 'Methods for estimating geometric parameters of retinal vessels using diagnostic images of fundus'. Proceedings of SPIE Conference on Optical Information Science and Computer and Holographic Optics and Image Processing, **3460**, pp. 316–325

JASIOBEDSKI, P., TAYLOR, C., and BRUNT, J. (1993): 'Automated analysis of retinal images', *Image Vision Comput.*, **11**, pp. 139–144

JAVITT, J., AIELLO, L., CHIANG, Y., FERRIS, F., CANNER, J., and GREENFIELD, S. (1994): 'Preventive eye care in people with diabetes is cost-saving to the federal government', *Diabetes Care*, **17**, pp. 909–917

KATZ, N., GOLDBAUM, M., NELSON, M., and CHAUDHURI, S. (1988): 'An image processing system for automatic retina diagnosis'. Proceedings of SPIE Conference on Three-Dimensional Imaging and Remote Sensing Imaging, pp. 131–137

KLEIN, R., KLEIN, B., MAGLI, Y., BROTHERS, R., MEUER, S., MOSS, S., and DAVIS, M. (1986): 'An alternative method of grading diabetic retinopathy', *Ophthalmology*, **93**, pp. 1183–1187

KLEIN, R., MOSS, S., KLEIN, B., DAVIS, M., and DEMETS, D. (1989): 'The Wisconsin epidemiologic study of diabetic retinopathy: XI The incidence of macular edema', *Ophthalmology*, **96**, pp. 1501–1510

KOCHNER, B., SCHULMAN, D., OBERMAIER, M., ZAHLMANN, G., MANN, G., and ENGLMEIER, K. H. (1997): 'An image processing system for analysing color fundus photographs with regard to diabetic retinopathy', *Klinische Monatsblatter fur Augenheilkunde*, **211**, p. 11

KOCHNER, B., SCHUHMANN, D., MICHAELIS, M., MANN, G., and ENGLEMEIER, K.-H. (1998): 'Course tracking and contour extraction of retinal vessels from color fundus photographs: most efficient use of steerable filters for model based image analysis'. Proceedings of SPIE Conference on Medical Imaging, pp. 755–761

KOZOUSEK, V., SHEN, Z., GREGSON, P., and SCOTT, R. (1992): 'Automated detection and quantification of venous beading using fourier analysis', *Can. J. Ophthalmol.*, **27**, pp. 288–294

KRISTINSSON, J., GOTTFREDSDOTTIR, M., and STEFANSSON, E. (1997): 'Retinal vessel dilation and elongation precedes diabetic macular oedema', *Br. J. Ophthalmol.*, **81**, pp. 274–278

LANDINI, G., MISSON, G., and MURRAY, P. (1993): 'Fractal analysis of the normal human retinal flourescin angiogram', *Current Eye Res.*, **12**, pp. 23–27

LANDINI, G., MURRAY, P., and MISSON, C. (1995): 'Local connected fractal dimensions and lacunarity analysis of 60 fluorecin angiograms', *Investigat. Ophthalmol. Visual Sci.*, **36**, pp. 2749–2755

LEE, S. (1992): 'Identifying retinal vessel in ocular fundus images'. PhD thesis, The University of New Mexico, USA

LEE, S., and WANG, Y. (1998): 'A general algorithm of recognizing small, vague and imager-alike objects in a nonuniformly illuminated medical diagnostic image'. Proceedings of 32nd Conference on Signals Systems and Computers, **2**, Asilomar, pp. 941–943

LEE, S., and WANG, Y. (1999): 'Automatic retinal image quality assessment and enhancement'. SPIE Conference on Image Processing, **3661**, pp. 1581–1590

LEE, S., WANG, Y., and LEE, E. (1999): 'A computer algorithm for automated detection and quantification of microaneurysms and hemorrhages (hma's) in color retinal images'. SPIE Conference on Image Perception and Performance, **3663**, pp. 61–71

LIESENFELD, B., KOHNER, E., PIEHLMEIER, W., KLUTHE, S., PORTA, M., BEK, T., OBERMAIER, M., MAYER, H., MANN, G., HOLLE, R., and HEP, K. (2000): 'A telemedical approach to the screening of diabetic retinopathy: Digital fundus photography', *Diabetes Care*, **23**, pp. 345–348

MARTÍNEZ-PÉREZ, M., HUGHES, A., STANTON, A., THOM, S., BHARATH, A., and PARKER, K. (1999a): 'Segmentation of retinal blood vessels based on the second directional derivative and region growing'. Proceedings of International Conference on Image Processing, New Jersey, USA, pp. 173–176

MARTÍNEZ-PÉREZ, M., HUGHES, A., STANTON, A., THOM, S., BHARATH, A., and PARKER, K. (1999b): 'Retinal blood vessel segmentation by means of scale-space analysis and region growing'. Proceedings of 2nd International Conference on Medical Image Computing and Computer Assisted Intervention, Cambridge, UK, pp. 90–97

MENDONÇA, A., CAMPILHO, A., and NUNES, J. (1999): 'Automatic segementation of microaneurysms in retinal angiograms of diabetic patients'. Proceedings of 10th International Conference on Image Analysis and Processing, Los Alamitos, pp. 728–723

NATIONAL SCREENING COMMITTEE (2000): 'Second report of the United Kingdom national screening committee'. Department of Health, PO Box 777, London SE1 6XH, United Kindom.

NGUYEN, H., BUTTLER, M., ROYCHOUDHRY, A., SHANNON, A., FLACK, J., and MITCHEL, P. (1996): 'Classification of diabetic retinopathy using neural networks'. Proceedings of 18th Annual International Conference of IEEE Engineering in Medicine and Biology Society, IEEE, Amsterdam, The Netherlands, Vol. 4, pp. 1548–1549

ØIEN, G., and OSNES, P. (1995): 'Diabetic retinopathy: Automatic detection of early symptoms from retinal images'. Proceedings of Norwegian Signal Processing Symposium, Stavanger, Norway, Vol. VII, pp. 135–140

OKAZAKI, K., and TAMURA, S. (1983): 'Spherical shading correction of eye fundus image'. Proceedings of International Conference on Systems, Man and Cybernetics, Bombay, India, pp. 1084–1087

OTSU, N. (1979): 'A threshold method from gray-level histograms', *IEEE Trans. Syst. Man Cybern.*, **9**, pp. 62–66

PANICO, J., and STERLING, P. (1995): 'Retinal neurons and vessels are not fractal but space-filling', *J. Compar. Neurol.*, **361**, pp. 479–490

PATEL, V. (1995): 'Diabetic retinopathy: haemodynamic and clinical factors in the pathogenesis' (Verlag, Josef Eul, Koln, 1995 (ISBN: 3-89012-432-1)

PELI, E. (1993): 'Enhancement of retinal images: Pros and problems', *Neurosci. Biobehav. Rev.*, **17**, pp. 477–482

PHILLIPS, R., SPENCER, T., ROSS, P., SHARP, P., and FORRESTER, J. (1991): 'Quantification of diabetic maculopathy by digital imaging of the fundus', *Eye*, **5**, pp. 130–137

POLI, R., and VALLI, G. (1997): 'An algorithm for real-time vessel enhancement and detection', *Comput. Methods Programs Biomed.*, **52**, pp. 1–22

RASSAM, S., PATEL, V., BRINCHMANN-HANSEN, O., ENGVOLD, O., and KOHNER, E. (1994): 'Accurate vessel width measurement from fundus photographs: a new concept', *Br. J. Ophthalmol.*, **78**, pp. 24–29

RYAN, N., and HENEGHAN, C. (1999): 'Image registration techniques for digital ophthalmic images'. Irish Signal and Systems Conference, Galway, Ireland, pp. 301–308

SCHALKOFF, R. (1989): '*Digital image processing and computer vision*' (Wiley and Sons, Inc., New York, 1989)

SCOTT, J., FISHER, R., and PALMER, E. (1996): 'Neural networks in ventilation-perfusion imaging; part ii. Effects of interpretive variability', *Radiology*, **198**, pp. 707–713

SHIN, D., KAISER, R., LEE, M., and BERGER, J. (1999): 'Fundus image change analysis:geometric and radiometric normalisation'. Proceedings of SPIE Conference on Ophthalmic Technologies, **3591**, San Jose, California, pp. 129–136

SINCLAIR, S., GUPTA, A., and BHASIN, S. (1996): 'Automated lesion detection and grading of retinopathy from fundus photographs of diabetics', *Diabetes*, **45**, p. 192A

SINTHANAYOTHIN, C., BOYCE, J., COOK, H., and WILLIAMSON, T. (1999): 'Automated localisation of the optic disk, fovea, and retinal blood vessels from digital colour fundus images', *Br. J. Ophthalmol.*, **83**, pp. 902–910

SKOVBORG, F., NIELSEN, A., LAURITZEN, E., and HARKTOPP, O. (1969): 'Diameters of the retinal vessels in diabetic and normal subjects', *Diabetes*, **18**, pp. 292–298

SOLIZ, P., NEMETH, S., SWIFT, M., EDWARDS, A., MEUER, S., and BERGER, J. (2000): 'Improving the visualisation of drusen in age-related macular degeneration through maximum entropy digitization and stereo viewing'. Proceedings of SPIE Conference on Medical Imaging 2000: Image Perception and Performance, **3981**, pp. 217–281

SPENCER, T., OLSON, J., SHARP, P., and FORRESTER, J. (1994): 'A 'region-growing' approach to the quantification of microaneurysms in the diabetic retina'. BES/IPSM Meeting Engineering and Technology in Diabetes, Bournemouth, United Kingdom, p. 12

SPENCER, T., OLSON, J., MCHARDY, K., SHARP, P., and FORRESTER, J. (1996): 'An image-processing strategy for the segmentation and quantification of microaneurysms in flourescein angiograms of the ocular fundus', *Comput. Biomed. Res.*, **29**, pp. 284–302

TAMURA, S., OKAMOTO, Y., and YANASHIMA, K. (1988): 'Zero-crossing interval correction in tracing eye-fundus blood vessels', *Pattern Recognit.*, **21**, pp. 227–233

TANAKA, M., and TANAKA, K. (1988): 'An automatic technique for fundus-photograph mosaic and vascular net reconstruction'. Proceedings of MEDINFO, pp. 116–120

THOMPSON, H. (1999): 'A wavelet-based method for automated computer detection of diabetic retinopathy in fundus photographs', *Invest. Opthal. Visual Science*, **40**, p. S121

TOLIAS, Y., and PANAS, S. (1998): 'A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering', *IEEE Trans. Med. Imaging*, **17**, pp. 263–273

WANG, Y., and LEE, S. (1997): 'A fast method for automated detection of blood vessels in retinal images'. Asilomar Conference, IEEE Computer Society, Pacific Grove, CA, Vol. 2, pp. 1700–1704

WANG, Y., TOONEN, H., and MEYER-EBRECHT, D. (1990): 'A new method of automatic tracking, measuring blood vessels in retinal image'. Annual International Conference of IEEE Engineering in Medicine and Biology Society, Vol. 12, Philadelphia, USA, pp. 174–175

WARD, N., TOMLINSON, S., and TAYLOR, C. (1988): 'Image analysis of fundus photographs', *Ophthalmology*, **96**, pp. 80–86

WEST, G., BROWN, J., and ENQUIST, B. (1997): 'A general model for the origin of allometric scalling laws in biology', *Science*, **276**, pp. 122–126

WOOD, S. (1990): 'Analysis of retinal vessel structure from multiple images'. Annual International Conference of IEEE Engineering in Medicine and Biology Society, Vol. 12, Philadelphia, pp. 176–177

YANG, C.-W., MA, D.-J., WANG, C.-M., WEN, C.-H., LO, C.-S., and CHANG, C.-I. (2000): 'Computer-aided diagnostic detection system of venous beading in retinal images', *Opt. Eng.*, **39**, pp. 1293–1303

YE, D., and ZHENG, L. (1995): 'Fundus image processing and feature classification based on mathematical morphology method'. Proceedings of the Canadian Medical and Biological Engineers Conference, Vol. 2, pp. 1015–1016

YOUNG, S., GEORGE, L., LUSTY, J., and OWENS, D. (1997): 'A new screening tool for diabetic retinopathy: the Canon CR5 45NM retinal camera with Frost Medical Software RIS-Lite digital imaging system', *J. Audiovisual Media Med.*, **20**, pp. 11–14

YU, J., HUNG, B., and SUN, H. (1990): 'Automatic recognition of retinopathy from retinal images'. Proceedings of Twelfth Annual International Conference on IEEE Engineering and Medicine and Biology Society, Philadelphia, Vol. 12, pp. 171–173

YULONG, M., and DINGRU, X. (1990): 'Recognizing the glaucoma from ocular fundus image by image processing'. Proceedings of Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Philadelphia, Vol. 12, pp. 178–179

ZAHLMANN, G., SCHUBERT, M., OBERMAIER, M., and MANN, G. (1996): 'Concept of a knowledge based monitoring system for glaucoma and diabetic retinopathy using a telemedicine approach'. 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Amsterdam, pp. 1230–1231

ZAHLMANN, G., KOCHNER, B., UGI, I., SCHULHMANN, D., LIESENFELD, B., WEGNER, A., OBERMAIER, M., and MERTZ, M. (2000): 'Hybrid fuzzy image processing for situation assessment: A knowledge-based system for early detection of diabetic retinopathy', *IEEE Eng. Med. Biol.*, **19**, pp. 76–83

ZHOU, L., RZESZOTARSKI, M., SINGERMAN, L., and CHOKREFF, J. (1994): 'The detection and quantification of retinopathy using digital angiograms', *IEEE Trans. Med. Imag.*, **13**, 619–626

Authors' biographies

THOMAS TENG received his BSc in Medical Instrumentation and Computing from Bournemouth University in 1997. He is currently undertaking research into the development and application of intelligent systems for medical image processing.

MARTIN LEFLEY is a reader in intelligent interfaces at Bournemouth University. He has a Masters in Statistics from Newcastle University and was awarded his PhD for utilising computer vision in knitted fabric inspection, in 1988. He is investigating the practical applications of artificial intelligence. Many of these projects used neural networks to realise computer visual abilities.

DENZIL CLAREMONT is Professor of Biomedical Engineering at Bournemouth University. He received a BSc honours degree in Physiology from the University of London in 1975, an MSc in Bioengineering from Strathclyde University in 1976 and a DPhil from Sussex University in 1981. His research interests include physiological measurements and in the development and application of engineering in the management of diabetes.

This page intentionally contains only this sentence.

Appendix B

Sample Morphology Experiment Data

In this section some data are given for the morphological and neural trials. Two sets of data from each methodlgy are provided as a sample. The following table presents the sensitivity and specificity of the initial 10 images on which the morphological algorithm was developed. It covers a range of 0 to 100 for parameter t in Equation 3.14 The subsequent table presents similar results but for the verification dataset.

Table B.1: True Positive *:* True Negative fractions for the RBH images used in development. A rolling ball radius of 15 and connected component size of 30 picture elements was used.

	Image i									
t	0	1	2	3	4	5	6	7	8	9
0	1.000:0.235	1.000:0.235	1.000:0.230	1.000:0.233	1.000:0.229	1.000:0.226	1.000:0.236	1.000:0.235	1.000:0.231	1.000:0.230
1	0.993:0.303	0.991:0.302	0.995:0.291	0.993:0.297	0.993:0.288	0.984:0.281	0.994:0.295	0.996:0.296	0.997:0.377	0.995:0.366
2	0.994:0.298	0.992:0.297	0.996:0.296	0.995:0.295	0.994:0.292	0.981:0.297	0.996:0.305	0.996:0.290	0.997:0.387	0.997:0.376
3	0.994:0.305	0.992:0.296	0.997:0.330	0.996:0.314	0.993:0.333	0.962:0.356	0.995:0.347	0.997:0.302	0.998:0.419	0.996:0.415
4	0.994:0.305	0.992:0.296	0.997:0.330	0.996:0.314	0.993:0.333	0.962:0.356	0.995:0.347	0.997:0.302	0.998:0.419	0.996:0.415
5	0.996:0.329	0.993:0.305	0.995:0.385	0.995:0.354	0.987:0.397	0.928:0.445	0.991:0.411	0.996:0.328	0.996:0.464	0.994:0.474
6	0.996:0.367	0.994:0.326	0.992:0.458	0.993:0.409	0.976:0.482	0.878:0.554	0.986:0.492	0.995:0.365	0.994:0.523	0.989:0.548
7	0.995:0.415	0.992:0.357	0.986:0.540	0.989:0.478	0.958:0.578	0.810:0.670	0.977:0.581	0.994:0.414	0.990:0.589	0.981:0.628
8	0.992:0.473	0.988:0.397	0.978:0.625	0.982:0.555	0.931:0.675	0.720:0.779	0.964:0.669	0.992:0.470	0.985:0.657	0.971:0.709
9	0.970:0.561	0.981:0.451	0.965:0.704	0.972:0.636	0.894:0.763	0.618:0.866	0.946:0.750	0.987:0.533	0.975:0.722	0.957:0.780
10	0.977:0.609	0.977:0.503	0.949:0.774	0.956:0.712	0.847:0.836	0.514:0.926	0.922:0.817	0.983:0.598	0.962:0.780	0.962:0.780
11	0.968:0.670	0.971:0.552	0.929:0.831	0.936:0.780	0.790:0.890	0.412:0.963	0.893:0.869	0.976:0.664	0.947:0.828	0.912:0.886
12	0.953:0.731	0.962:0.609	0.907:0.876	0.912:0.836	0.723:0.928	0.316:0.983	0.862:0.907	0.968:0.724	0.926:0.867	0.880:0.921

					Image					
t	0	1	2	3	4	5	6	7	8	9
13	0.936:0.785	0.954:0.665	0.879:0.910	0.879:0.881	0.655:0.953	0.231:0.992	0.825:0.935	0.957:0.779	0.902:0.897	0.843:0.944
14	0.915:0.831	0.942:0.718	0.846:0.935	0.842:0.914	0.584:0.969	0.165:0.996	0.786:0.953	0.943:0.826	0.875:0.921	0.803:0.961
15	0.889:0.867	0.928:0.766	0.813:0.953	0.801:0.939	0.517:0.979	0.112:0.998	0.744:0.965	0.926:0.865	0.847:0.939	0.762:0.972
16	0.865:0.894	0.915:0.810	0.778:0.966	0.754:0.956	0.452:0.985	0.077:0.999	0.701:0.974	0.907:0.896	0.815:0.953	0.720:0.979
17	0.835:0.917	0.898:0.847	0.741:0.975	0.702:0.968	0.390:0.990	0.043:0.999	0.656:0.980	0.884:0.919	0.781:0.964	0.677:0.984
18	0.801:0.934	0.877:0.877	0.703:0.981	0.652:0.976	0.332:0.993	0.024:0.999	0.611:0.984	0.858:0.937	0.751:0.971	0.636:0.988
19	0.764:0.948	0.855:0.901	0.669:0.986	0.604:0.981	0.281:0.995	0.013:0.999	0.566:0.988	0.833:0.950	0.721:0.977	0.597:0.991
20	0.728:0.958	0.834:0.921	0.634:0.989	0.555:0.985	0.228:0.996	0.005:1.000	0.526:0.990	0.808:0.961	0.687:0.982	0.558:0.993
21	0.693:0.966	0.811:0.936	0.601:0.992	0.510:0.988	0.186:0.997	0.002:1.000	0.486:0.992	0.781:0.968	0.659:0.985	0.520:0.994
22	0.655:0.972	0.785:0.948	0.565:0.993	0.465:0.991	0.149:0.998	0.001:1.000	0.450:0.993	0.750:0.974	0.631:0.988	0.483:0.995
23	0.615:0.977	0.757:0.957	0.534:0.995	0.424:0.992	0.117:0.999	0.000:1.000	0.415:0.995	0.723:0.979	0.604:0.991	0.445:0.996
24	0.580:0.981	0.732:0.964	0.502:0.996	0.386:0.994	0.090:0.999	0.000:1.000	0.385:0.996	0.692:0.983	0.579:0.992	0.413:0.997
25	0.545:0.984	0.707:0.969	0.471:0.997	0.349:0.995	0.065:0.999	0.000:1.000	0.356:0.996	0.663:0.986	0.557:0.994	0.377:0.998
26	0.514:0.987	0.682:0.973	0.444:0.997	0.316:0.996	0.050:1.000	0.000:1.000	0.328:0.997	0.633:0.988	0.533:0.995	0.343:0.998

Table B.1: (continued from previous page)

	Image									
t	0	1	2	3	4	5	6	7	8	9
27	0.480:0.989	0.659:0.977	0.417:0.998	0.288:0.997	0.038:1.000	0.000:1.000	0.301:0.997	0.602:0.990	0.511:0.995	0.311:0.999
28	0.450:0.991	0.637:0.980	0.393:0.998	0.263:0.997	0.029:1.000	0.000:1.000	0.278:0.998	0.572:0.992	0.486:0.996	0.280:0.999
29	0.420:0.992	0.615:0.982	0.371:0.998	0.240:0.998	0.021:1.000	0.000:1.000	0.257:0.998	0.541:0.993	0.464:0.997	0.254:0.999
30	0.390:0.993	0.594:0.984	0.351:0.999	0.221:0.998	0.016:1.000	0.000:1.000	0.237:0.998	0.513:0.994	0.442:0.997	0.233:0.999
31	0.363:0.994	0.575:0.986	0.332:0.999	0.204:0.999	0.012:1.000	0.000:1.000	0.217:0.999	0.484:0.995	0.419:0.998	0.212:0.999
32	0.335:0.995	0.557:0.987	0.314:0.999	0.188:0.999	0.009:1.000	0.000:1.000	0.198:0.999	0.457:0.996	0.397:0.998	0.191:1.000
33	0.309:0.996	0.540:0.989	0.295:0.999	0.174:0.999	0.006:1.000	0.000:1.000	0.182:0.999	0.429:0.997	0.376:0.998	0.172:1.000
34	0.286:0.997	0.521:0.990	0.279:0.999	0.161:0.999	0.004:1.000	0.000:1.000	0.168:0.999	0.400:0.997	0.353:0.999	0.155:1.000
35	0.265:0.997	0.502:0.991	0.261:0.999	0.150:0.999	0.003:1.000	0.000:1.000	0.154:0.999	0.373:0.998	0.330:0.999	0.140:1.000
36	0.245:0.998	0.485:0.992	0.246:0.999	0.138:1.000	0.002:1.000	0.000:1.000	0.141:0.999	0.345:0.998	0.307:0.999	0.122:1.000
37	0.226:0.998	0.468:0.993	0.231:0.999	0.129:1.000	0.001:1.000	0.000:1.000	0.127:0.999	0.319:0.998	0.286:0.999	0.110:1.000
38	0.209:0.998	0.450:0.993	0.217:1.000	0.121:1.000	0.001:1.000	0.000:1.000	0.115:0.999	0.295:0.999	0.267:0.999	0.097:1.000
39	0.193:0.999	0.434:0.994	0.202:0.999	0.112:1.000	0.001:1.000	0.000:1.000	0.104:1.000	0.270:0.999	0.247:0.999	0.083:1.000
40	0.179:0.999	0.418:0.995	0.188:1.000	0.103:1.000	0.001:1.000	0.000:1.000	0.094:1.000	0.246:0.999	0.229:1.000	0.073:1.000

	Image									
t	0	1	2	3	4	5	6	7	8	9
41	0.165:0.999	0.402:0.995	0.176:1.000	0.095:1.000	0.001:1.000	0.000:1.000	0.084:1.000	0.224:0.999	0.212:1.000	0.063:1.000
42	0.152:0.999	0.386:0.996	0.165:1.000	0.087:1.000	0.000:1.000	0.000:1.000	0.075:1.000	0.204:0.999	0.194:1.000	0.056:1.000
43	0.141:0.999	0.371:0.996	0.154:1.000	0.078:1.000	0.000:1.000	0.000:1.000	0.067:1.000	0.185:0.999	0.179:1.000	0.048:1.000
44	0.131:0.999	0.358:0.996	0.143:1.000	0.073:1.000	0.000:1.000	0.000:1.000	0.062:1.000	0.166:1.000	0.163:1.000	0.041:1.000
45	0.121:0.999	0.343:0.997	0.131:1.000	0.067:1.000	0.000:1.000	0.000:1.000	0.056:1.000	0.150:1.000	0.150:1.000	0.035:1.000
46	0.112:1.000	0.331:0.997	0.123:1.000	0.062:1.000	0.000:1.000	0.000:1.000	0.051:1.000	0.134:1.000	0.137:1.000	0.031:1.000
47	0.104:1.000	0.318:0.997	0.113:1.000	0.056:1.000	0.000:1.000	0.000:1.000	0.046:1.000	0.121:1.000	0.125:1.000	0.027:1.000
48	0.096:1.000	0.303:0.998	0.102:1.000	0.050:1.000	0.000:1.000	0.000:1.000	0.042:1.000	0.109:1.000	0.114:1.000	0.023:1.000
49	0.088:1.000	0.290:0.998	0.093:1.000	0.046:1.000	0.000:1.000	0.000:1.000	0.038:1.000	0.095:1.000	0.104:1.000	0.021:1.000
50	0.082:1.000	0.277:0.998	0.085:1.000	0.041:1.000	0.000:1.000	0.000:1.000	0.033:1.000	0.083:1.000	0.094:1.000	0.018:1.000
51	0.075:1.000	0.264:0.998	0.078:1.000	0.036:1.000	0.000:1.000	0.000:1.000	0.029:1.000	0.073:1.000	0.086:1.000	0.016:1.000
52	0.069:1.000	0.252:0.998	0.072:1.000	0.033:1.000	0.000:1.000	0.000:1.000	0.026:1.000	0.063:1.000	0.078:1.000	0.013:1.000
53	0.063:1.000	0.240:0.999	0.066:1.000	0.030:1.000	0.000:1.000	0.000:1.000	0.023:1.000	0.054:1.000	0.070:1.000	0.012:1.000
54	0.058:1.000	0.229:0.999	0.062:1.000	0.026:1.000	0.000:1.000	0.000:1.000	0.020:1.000	0.045:1.000	0.063:1.000	0.010:1.000

Table B.1: (continued from previous page)

	Image									
t	0	1	2	3	4	5	6	7	8	9
55	0.053:1.000	0.216:0.999	0.059:1.000	0.023:1.000	0.000:1.000	0.000:1.000	0.018:1.000	0.038:1.000	0.055:1.000	0.008:1.000
56	0.048:1.000	0.205:0.999	0.055:1.000	0.021:1.000	0.000:1.000	0.000:1.000	0.015:1.000	0.032:1.000	0.049:1.000	0.007:1.000
57	0.045:1.000	0.194:0.999	0.194:0.999	0.052:1.000	0.018:1.000	0.000:1.000	0.000:1.000	0.013:1.000	0.026:1.000	0.042:1.000
58	0.042:1.000	0.183:0.999	0.049:1.000	0.016:1.000	0.000:1.000	0.000:1.000	0.011:1.000	0.021:1.000	0.037:1.000	0.004:1.000
59	0.038:1.000	0.174:0.999	0.046:1.000	0.015:1.000	0.000:1.000	0.000:1.000	0.009:1.000	0.017:1.000	0.030:1.000	0.004:1.000
60	0.034:1.000	0.164:0.999	0.043:1.000	0.013:1.000	0.000:1.000	0.000:1.000	0.008:1.000	0.013:1.000	0.025:1.000	0.002:1.000
61	0.032:1.000	0.153:0.999	0.041:1.000	0.012:1.000	0.000:1.000	0.000:1.000	0.007:1.000	0.011:1.000	0.022:1.000	0.002:1.000
62	0.029:1.000	0.144:0.999	0.038:1.000	0.011:1.000	0.000:1.000	0.000:1.000	0.006:1.000	0.008:1.000	0.018:1.000	0.001:1.000
63	0.026:1.000	0.135:1.000	0.034:1.000	0.010:1.000	0.000:1.000	0.000:1.000	0.005:1.000	0.006:1.000	0.015:1.000	0.000:1.000
64	0.024:1.000	0.126:1.000	0.032:1.000	0.009:1.000	0.000:1.000	0.000:1.000	0.004:1.000	0.005:1.000	0.012:1.000	0.000:1.000
65	0.022:1.000	0.117:1.000	0.029:1.000	0.008:1.000	0.000:1.000	0.000:1.000	0.004:1.000	0.004:1.000	0.010:1.000	0.000:1.000
66	0.020:1.000	0.109:1.000	0.027:1.000	0.007:1.000	0.000:1.000	0.000:1.000	0.003:1.000	0.003:1.000	0.008:1.000	0.000:1.000
67	0.018:1.000	0.102:1.000	0.026:1.000	0.006:1.000	0.000:1.000	0.000:1.000	0.002:1.000	0.002:1.000	0.006:1.000	0.000:1.000
68	0.016:1.000	0.095:1.000	0.023:1.000	0.005:1.000	0.000:1.000	0.000:1.000	0.002:1.000	0.002:1.000	0.005:1.000	0.000:1.000

continued on next page

Table B.1: (continued from previous page)

	Image									
t	0	1	2	3	4	5	6	7	8	9
69	0.014:1.000	0.090:1.000	0.021:1.000	0.005:1.000	0.000:1.000	0.000:1.000	0.001:1.000	0.001:1.000	0.004:1.000	0.000:1.000
70	0.012:1.000	0.085:1.000	0.020:1.000	0.005:1.000	0.000:1.000	0.000:1.000	0.001:1.000	0.001:1.000	0.003:1.000	0.000:1.000
71	0.011:1.000	0.078:1.000	0.018:1.000	0.004:1.000	0.000:1.000	0.000:1.000	0.001:1.000	0.000:1.000	0.002:1.000	0.000:1.000
72	0.009:1.000	0.073:1.000	0.017:1.000	0.003:1.000	0.000:1.000	0.000:1.000	0.001:1.000	0.000:1.000	0.001:1.000	0.000:1.000
73	0.007:1.000	0.068:1.000	0.015:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.001:1.000	0.000:1.000	0.001:1.000	0.000:1.000
74	0.007:1.000	0.063:1.000	0.014:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.001:1.000	0.000:1.000
75	0.006:1.000	0.058:1.000	0.012:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
76	0.004:1.000	0.055:1.000	0.011:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
77	0.004:1.000	0.051:1.000	0.010:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
78	0.003:1.000	0.047:1.000	0.009:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
79	0.003:1.000	0.044:1.000	0.008:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
80	0.003:1.000	0.041:1.000	0.007:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
81	0.002:1.000	0.038:1.000	0.006:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
82	0.002:1.000	0.035:1.000	0.006:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000

Table B.1: (continued from previous page)

	Image									
t	0	1	2	3	4	5	6	7	8	9
83	0.001:1.000	0.033:1.000	0.005:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
84	0.001:1.000	0.030:1.000	0.004:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
85	0.001:1.000	0.028:1.000	0.004:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
86	0.001:1.000	0.026:1.000	0.003:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
87	0.000:1.000	0.024:1.000	0.024:1.000	0.003:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
88	0.000:1.000	0.022:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
89	0.000:1.000	0.020:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
90	0.000:1.000	0.019:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
91	0.000:1.000	0.018:1.000	0.002:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
92	0.000:1.000	0.016:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
93	0.000:1.000	0.015:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
94	0.000:1.000	0.014:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
95	0.000:1.000	0.012:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
96	0.000:1.000	0.011:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000

Table B.1: (continued from previous page)

t	0	1	2	3	4	5	6	7	8	9
					Image					
97	0.000:1.000	0.010:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
98	0.000:1.000	0.009:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
99	0.000:1.000	0.008:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000

Table B.2:   True Positive : True Negative fractions at Threhold value t, for the RBH data used for verification (images 10–14 from Table 3.2. A rolling ball radius of 15 and connected component size of 30 picture elements was used.

	Image				
t	10	11	12	13	14
0	1.000:0.230	1.000:0.231	1.000:0.234	1.000:0.229	1.000:0.233
1	0.994:0.363	0.991:0.295	0.996:0.301	0.992:0.293	0.993:0.295
2	0.995:0.376	0.994:0.300	0.997:0.291	0.992:0.289	0.995:0.301
3	0.995:0.414	0.993:0.336	0.997:0.303	0.994:0.287	0.995:0.340
4	0.994:0.469	0.989:0.398	0.996:0.336	0.994:0.294	0.991:0.404
5	0.991:0.542	0.981:0.477	0.993:0.386	0.992:0.311	0.985:0.484
6	0.991:0.542	0.981:0.477	0.993:0.386	0.992:0.311	0.985:0.484
7	0.988:0.612	0.965:0.569	0.988:0.447	0.989:0.335	0.974:0.576
8	0.982:0.686	0.952:0.657	0.981:0.515	0.985:0.367	0.959:0.670
9	0.972:0.751	0.928:0.741	0.972:0.588	0.980:0.405	0.937:0.754
10	0.959:0.807	0.899:0.813	0.958:0.665	0.973:0.448	0.907:0.823
11	0.941:0.851	0.865:0.868	0.937:0.739	0.963:0.495	0.872:0.876
12	0.920:0.885	0.826:0.908	0.911:0.803	0.949:0.546	0.833:0.914
13	0.893:0.912	0.783:0.936	0.881:0.855	0.934:0.598	0.788:0.940
14	0.861:0.932	0.740:0.955	0.849:0.896	0.916:0.650	0.739:0.957
15	0.828:0.947	0.694:0.967	0.809:0.927	0.891:0.700	0.690:0.969
16	0.785:0.959	0.649:0.976	0.765:0.950	0.865:0.747	0.642:0.977
17	0.742:0.967	0.604:0.982	0.721:0.965	0.834:0.790	0.595:0.982
18	0.699:0.974	0.566:0.986	0.676:0.975	0.797:0.828	0.547:0.986
19	0.655:0.979	0.526:0.989	0.629:0.982	0.759:0.861	0.503:0.989
20	0.609:0.983	0.488:0.992	0.583:0.987	0.720:0.890	0.457:0.991

continued on next page

Table B.2: (continued from previous page)

			Image		
t	10	11	12	13	14
21	0.564:0.986	0.454:0.993	0.542:0.990	0.680:0.913	0.417:0.993
22	0.521:0.988	0.424:0.995	0.501:0.992	0.640:0.931	0.377:0.995
23	0.475:0.990	0.392:0.996	0.459:0.994	0.596:0.945	0.340:0.996
24	0.431:0.992	0.363:0.996	0.421:0.995	0.558:0.957	0.304:0.997
25	0.394:0.993	0.338:0.997	0.385:0.997	0.519:0.966	0.269:0.997
26	0.357:0.994	0.313:0.998	0.352:0.997	0.484:0.973	0.235:0.998
27	0.325:0.995	0.289:0.998	0.323:0.998	0.446:0.978	0.199:0.998
28	0.298:0.996	0.265:0.998	0.296:0.998	0.415:0.983	0.173:0.999
29	0.274:0.997	0.244:0.999	0.271:0.999	0.383:0.986	0.146:0.999
30	0.252:0.997	0.221:0.999	0.246:0.999	0.354:0.989	0.122:0.999
31	0.232:0.997	0.199:0.999	0.226:0.999	0.328:0.991	0.098:0.999
32	0.211:0.998	0.180:0.999	0.204:0.999	0.307:0.992	0.078:0.999
33	0.194:0.998	0.163:0.999	0.185:0.999	0.286:0.994	0.061:1.000
34	0.173:0.998	0.147:1.000	0.168:1.000	0.265:0.995	0.047:1.000
35	0.157:0.999	0.132:1.000	0.153:1.000	0.245:0.996	0.037:1.000
36	0.141:0.999	0.120:1.000	0.138:1.000	0.228:0.996	0.027:1.000
37	0.127:0.999	0.109:1.000	0.123:1.000	0.209:0.997	0.021:1.000
38	0.113:0.999	0.099:1.000	0.109:1.000	0.195:0.997	0.015:1.000
39	0.100:0.999	0.092:1.000	0.099:1.000	0.181:0.998	0.012:1.000
40	0.089:0.999	0.085:1.000	0.090:1.000	0.168:0.998	0.009:1.000
41	0.079:1.000	0.077:1.000	0.082:1.000	0.155:0.998	0.007:1.000
42	0.069:1.000	0.070:1.000	0.075:1.000	0.143:0.999	0.005:1.000
43	0.059:1.000	0.063:1.000	0.067:1.000	0.129:0.999	0.003:1.000
44	0.048:1.000	0.058:1.000	0.060:1.000	0.117:0.999	0.002:1.000

Table B.2: (continued from previous page)

	Image				
t	10	11	12	13	14
45	0.039:1.000	0.054:1.000	0.054:1.000	0.106:0.999	0.002:1.000
46	0.032:1.000	0.051:1.000	0.047:1.000	0.096:0.999	0.001:1.000
47	0.026:1.000	0.046:1.000	0.043:1.000	0.086:0.999	0.001:1.000
48	0.020:1.000	0.043:1.000	0.038:1.000	0.078:0.999	0.000:1.000
49	0.015:1.000	0.039:1.000	0.035:1.000	0.068:0.999	0.000:1.000
50	0.012:1.000	0.035:1.000	0.031:1.000	0.059:0.999	0.000:1.000
51	0.009:1.000	0.032:1.000	0.027:1.000	0.051:0.999	0.000:1.000
52	0.007:1.000	0.029:1.000	0.023:1.000	0.045:0.999	0.000:1.000
53	0.005:1.000	0.026:1.000	0.020:1.000	0.039:0.999	0.000:1.000
54	0.004:1.000	0.024:1.000	0.017:1.000	0.033:1.000	0.000:1.000
55	0.003:1.000	0.022:1.000	0.015:1.000	0.029:1.000	0.000:1.000
56	0.002:1.000	0.020:1.000	0.012:1.000	0.024:1.000	0.000:1.000
57	0.002:1.000	0.019:1.000	0.010:1.000	0.020:1.000	0.000:1.000
58	0.001:1.000	0.017:1.000	0.008:1.000	0.017:1.000	0.000:1.000
59	0.001:1.000	0.015:1.000	0.007:1.000	0.015:1.000	0.000:1.000
60	0.000:1.000	0.014:1.000	0.006:1.000	0.012:1.000	0.000:1.000
61	0.000:1.000	0.012:1.000	0.005:1.000	0.010:1.000	0.000:1.000
62	0.000:1.000	0.011:1.000	0.004:1.000	0.008:1.000	0.000:1.000
63	0.000:1.000	0.009:1.000	0.003:1.000	0.007:1.000	0.000:1.000
64	0.000:1.000	0.009:1.000	0.003:1.000	0.006:1.000	0.000:1.000
65	0.000:1.000	0.008:1.000	0.002:1.000	0.004:1.000	0.000:1.000
66	0.000:1.000	0.007:1.000	0.002:1.000	0.003:1.000	0.000:1.000
67	0.000:1.000	0.006:1.000	0.002:1.000	0.003:1.000	0.000:1.000
68	0.000:1.000	0.005:1.000	0.002:1.000	0.002:1.000	0.000:1.000

Table B.2: (continued from previous page)

			Image		
t	10	11	12	13	14
69	0.000:1.000	0.004:1.000	0.002:1.000	0.002:1.000	0.000:1.000
70	0.000:1.000	0.003:1.000	0.002:1.000	0.001:1.000	0.000:1.000
71	0.000:1.000	0.003:1.000	0.001:1.000	0.001:1.000	0.000:1.000
72	0.000:1.000	0.002:1.000	0.001:1.000	0.001:1.000	0.000:1.000
73	0.000:1.000	0.002:1.000	0.001:1.000	0.001:1.000	0.000:1.000
74	0.000:1.000	0.001:1.000	0.000:1.000	0.001:1.000	0.000:1.000
75	0.000:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000
76	0.000:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000
77	0.000:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000
78	0.000:1.000	0.001:1.000	0.000:1.000	0.000:1.000	0.000:1.000
79	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
80	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
81	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
82	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
83	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
84	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
85	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
86	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
87	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
88	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
89	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
90	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
91	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
92	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000

Table B.2: (continued from previous page)

	Image				
t	10	11	12	13	14
93	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
94	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
95	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
96	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
97	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
98	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000
99	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000	0.000:1.000

References

Adler, F. (1992). Adler's physiology of the eye: clinical application.,
   In: W. Hart (editor), *Adler's Physiology of the Eye.*, 9th edn, St Louis,
   London. ISBN: 0801621070 Shelfmark: (B) FD 40 Author: Adler
   Francis Heed 1895- Title: Adler's Physiology of the eye clinical ap-
   plication Edition: 9th ed. edited by William M. Hart, Jr. Publisher:
   St Louis, Mo. Mosby 1992.

Aiello, L., Bursell, S., Cavallerano, J., Kelly Gardner, W. and Strong, J.
   (1998a). Joslin vision network validation study: Pilot image sta-
   bilization phase, *Journal of the American Optometric Association*
   69(11): 699–710.

Aiello, L., Cavallerano, J., Gardner, T., King, G., Blankenship, G., Ferris
   3rd, F. and Klein, R. (1998b). Diabetic retinopathy, *Diabetes Care*
   21(1): 143–156.

Akita, K. and Kuga, H. (1979). Towards understanding color ocular fundus
   images, In: *Proc. of the 6th IJCAI*, Tokio, Japan, pp. 7–12.

Akita, K. and Kuga, H. (1980). Digital processing of color ocular fundus
   images, In: D. Lindberg and S. Kaihara (editors), *Proc. MEDINFO*,
   Tokyo, pp. 80–84.

Akita, K. and Kuga, H. (1982). A computer method of understanding ocu-
   lar fundus images, *Pattern Recognition* 16(6): 431–443.

Aldington, S., Kohner, E., Meuer, S., Klein, R. and Sjole, A. (1995).
Methodology for retinal photography and assessment of diabetic
retinopathy: the eurodiab iddm complications study., *Diabetologia*
38(4): 437–44.

American Academy Ophthalmology Prefered Practice Patterns Committee
Retina Panel (1998). Preferred practice pattern diabetic retinopathy,
Technical Report 110020, American Academy of Ophthalmology.

American Diabetes Association (1999). American diabetes association:
Clinical practice recomendations 1999, *Diabetes Care* 22(S1): S70–S73.

Baumgart-Schmitt, R., Hermann, W., Eilers, R. and Bes, F. (1997). On
the use of neural network techniques to analyse sleep EEG data, *Neu-
ropsycobiology* 36(4): 194–210.

Baxt, W. G. (1991a). Decision making and acute MI's [editorial; see com-
ments], *Ann Emerg Med* 20(11): 1266–7.

Baxt, W. G. (1991b). Use of an artificial neural network for the diagnosis
of myocardial infarction [published erratum appears in ann intern med
1992 jan 1;116(1):94] [see comments], *Ann Intern Med* 115(11): 843–8.
University of California, San Diego Medical Center.

Baxt, W. G. (1994). A neural network trained to identify the presence of
myocardial infarction bases some decisions on clinical associations that
differ from accepted clinical teaching, *Med Decis Making* 14(3): 217–
222.

Beach, J., Tiedeman, J., M., H. and Y., S. (1999). Multi-spectral fundus
imaging fro early detection of diabetic retinopathy, In: *Proc. SPIE
Conference on Clinical diagnostic Systems and Technologies*, Vol.
3603, San Jose, CA. USA, pp. 114–121.

Bhattacharya, U., Chaudhuri, B. and Paui, S. (1997). An mlp-based texture segmentation method without selecting a feature set., *Image and Vision Computing* 15(12): 937–948.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford. ISBN 0-19-853864-2 (Pbk).

Brinchmann-Hansen, O. and Engvold, O. (1986). Microphotometry of the blood column and the light streak on retinal vessels in fundus photographs., *Acta Ophthalmologica* 64(Supplement 179): 9–19.

Can, A., Shen, H., Turner, J., Tanenbaum, H. and Roysam, B. (1999). Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms., *IEEE Trans. Inform. Technol. Biomed.* 3(2): 125–138.

Chaudhuri, S., Chatterjee, S., Katz, N. and Goldbaum, M. (1989a). Automatic detection of the optic nerve in retinal images, In: *Proc. IEEE International Conf. on Image Processing*, Singapore, pp. 1–5.

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

Chen, H., Patel, V., Wiek, J., Rassam, S. and Kohner, E. (1994). Vessel diameter changes during the cardiac cycle, *Eye* 8: 97–103.

Chutatape, O., Zheng, L. and Krishnan, S. (1998). Retinal blood vessel detection and tracking by matched gaussian and kalman filters, In: *Proc. 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 20, Hong Kong, October, pp. 3144–3149.

Cree, M., Olson, J., McHardy, K., Sharp, P. and Forrester, J. (1997). A fully automated comparitive microaneurism digital detection system, *Eye* 11(5): 622–28.

Dasbach, E., Fryback, D., Newcomb, P., Klein, R. and Klein, B. (1991). Cost-effectiveness of strategies for detecting diabetic retinopathy, *Medical Care* 29(1): 20–39.

Davis, M., Fisher, M., Gangnon, R., Barton, F., Aiello, L., Chew, E., Ferris 3rd, F. and Knatterud, G. (1998). Risk factors for high-risk proliferative diabetic retinopathy and severe visual loss: Early treatment diabetic retinopathy study report #18, *Invest Ophthalmol Vis Sci* 39(2): 233–252.

Daxer, A. (1993a). Characterization of the neovascularisation process in diabetic retinopathy by means of fractal geometry: diagnostic implications, *Graefe's Arch. Clin. Exp. Ophthalmol.* 231(12): 681–686.

Daxer, A. (1993b). The fractal geometry of proliferative diabetic retinopathy: implications for the diagnosis and the process of retinal vasculogenesis., *Current Eye Research* 12(12): 1103–1109.

Delori, F. (1988). Noninvasive technique for oximetry of blood in retinal vessels, *Applied Optics* 27(6): 1113–1125.

Dougherty, S., Bowyer, K. W. and Kranenbury, C. (1998). Roc curve evaluation of edge detector performance, In: *Proc. IEEE International Conf. on Image Processing*, Vol. 2, Chicago,IL, Oct, pp. 525–529.

DRS Research Group (1981). Report vii: A modification of the airlie house classifcation of diabetic retinopathy, *Invest. Ophthalmol. Vis. Sci.* 21(1): 210–226. Refered to in [Klein 89-Wisc IDDM study] as part of the protocol for examination, recording and grading.

DRS Research Group (1987). Indications for photocoagulation treatment of diabetic retinopathy: Diabetic Retinopathy Study Report No. 14, *International Ophthalmology Clinics* 27(4): 239–253.

Dumskyj, M., Aldington, S., Dore, C. and Kohner, E. (1996). The accurate assessment of changes in retinal vessel diameter using multiple frame electorcardiograph synchronized fundus photography., *Current Eye Research* 15(6): 625–632.

Eberhart, R., Dobbins, R. and Hutton, L. (1991). Neural network paradigm comparisons for appendicitis diagnosis., In: *Proc-4th IEEE Symp. on Computer Based Medical Systems*, Baltimore, pp. 298–304.

Egan, J. (1975). *Signal Detection Theory and ROC Analysis*, Series in Cognition and Perception, Academic Press INC.

Ege, B., Hejlesen, O., Larsen, O. V., Møller, Jennings, B., Kerr, D. and Cavan, D. (2000). Screening for diabetic retinopathy using computer based image analysis and statistical classification, *Computer Methods and Programms in Biomedicine* 62(3): 165–175.

Egmont-Petersen, M., de Ridder, D. and Handels, H. (2002). Image processing with neural networks – a review, *Pattern Recognition* 35(10): 2279–2301.

Englmeier, K., S., B., Schmid, K., Maurino, M., Porta, M., Bek, T., Ege, B., Larsen, O. and Hejlesen, O. (2002). Multi-resolution retinal vessel tracker based on directional smoothing, 3(21): 230–237.

ETDRS Research Group (1985). Photocoagulation for diabetic macular edema: Early treatment diabetic retinopathy study report number 1., *Arch. Ophthalmol.* 103(12): 1796–1 806.

ETDRS Research Group (1991a). Early Treatment Diabetic Retinopathy Study Design and Baseline Patient Characteristics: ETDRS Report Number 7., *Ophthalmology* 98(5): 741–56.

ETDRS Research Group (1991b). Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12., *Ophthalmology* 98(5): 823–33.

ETDRS Research Group (1991c). Grading diabetic retinopathy from stereoscopic color fundus photographs- an extension of the modified airlie house classification: ETDRS report number 10., *Ophthalmology* 98(5): 786–806.

Feman, S., Leonardmartin, T., Andrews, J., Armbuster, C., Burdge, T., Debelak, J., Lanier, A. and Fisher, A. (1995). A quantative system to evaluate diabetic-retinopathy from fundus photographs, *Investigative Ophthalmology and Visual Science.* 36(1): 174–181.

Flewelling, R. (1995). Noninvasive optical monitoring, In: J. Bronzino (editor), *The Biomedical Engineering Handbook*, CRC Press, pp. 1346–1356.

Forrester, J. and Knott, R. (1997). Pathogenesis of diabetic retinopathy and cataract, In: J. Pickup and W. G. (editors), *Textbook of Diabetes*, 2nd edn, Vol. 2, Blackwell Science Ltd., London, pp. 4501–4518.

Frame, A., Undrill, P., Olson, J., McHardy, K., Sharp, P. and Forrester, J. (1997a). Structural analysis of retinal vessels, In: *Proc. 6th International Conf. on Image Processing and its Applications*, Vol. 2, IEE, Dublin, pp. 824–827.

Frame, A., Undrill, P., Olson, J., McHardy, K., Sharp, P. and Forrester, J. (1997b). Texture analysis of retinal neovascularisation., In: *Proc. of IEE Colloquium on pattern recognition*, Vol. 6, IEE, London, pp. 1–6.

Fransen, S., Leonard-Martin, T., Feuer, W., Hildebrand, P. and The Inoveon Health Research Group (2002). Clinical evaluation of patients with diabetic retinopathy, *Ophthalmology* 109(3): 595–601.

Gagnon, L., Lalonde, M., Beaulieu, M. and Boucher, M. (2001). Procedure to detect anatomical structures in optical fundus images, In: *Proc. SPIE Conf. on Medical Imaging.*, Vol. 2, San Diego, California, pp. 1218–1225.

Gang, L., Chutatape, O. and Krishnan, S. (2002). Detection and measurement of retinal vesels in fundus images using amplitude modified second-order gaussian filter, *IEEE Trans. on Biomedical Engineering* 49(2): 168–172.

Gao, X., Bharath, A. A., Hughes, A., A., S., Chapman, N. and Thom, S. (1997). Towards retinal vessel parameterisation, In: *Proc. SPIE Conf. on Medical Imaging:Image Processing*, Vol. 3034, pp. 734–744.

Gardner, G., Keating, D., Williamson, T. and Elliot, A. (1996). Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool, *British Journal of Ophthalmology* 80(11): 940–944.

George, L., Leverton, C., Young, S., Lusty, J., Dunstan, F. and Owens, D. (1997). Can digitised colour 35mm transparencies be used to diagnose diabetic retinopathy ?, *Diabetic Medicine* 14(11): 970–973.

Goh, K., Sarkodie-Gyan, T., Campbell, A., Simpson, D. and McNeela, B. (1997). Computer assisted photocoagulation for treatement of diabetic retinopathy:preliminary results., In: *Proc. 30th International Symp. on Automotive Technology and Automation.*, Vol. 2, Florence, pp. 655–662.

Goldbaum, M., Katz, N., Chaudhuri, S. and Nelson, M. (1989). Image understanding for automated retinal diagnosis., In: *Proc-IEEE Symp For Computer Aplications in Clinical Medicine.*, pp. 756–760.

Goldbaum, M., Kouznetsova, V., Cote, B., Hart, W. and Nelson, M. (1993). Automated registration of digital ocular fundus images for comparison of lesions, *SPIE: Ophthalmic Technologies III* 1877: 94–99.

Gonzalez, R. and Woods, R. (2002). *Digital Image Processing*, 2nd edn, Prentice Hall.

Gregson, P., Shen, Z., Scott, R. and Kozoudek, V. (1995). Automated grading of venous beading, *Computers and Biomedical Research* 28(4): 291.

Hammer, M., Leistritz, S., Leistritz, L., Schweitser, D., Thamm, E. and Donnerhacke, K, H. (1997). Monte-carlo simulation of retinal vessel profiles for the interpretation of in vivo oxymetric measurements by imaging fundus reflectometry, In: *Proc. of Medical Applications of Lasers In Dermatology, Ophthalmology, Dentistry, and Endoscopy,*, Vol. 3192, San Remo, Italy, pp. 211–218.

Hanley, J. and McNeil, B. (1992). The meaning and use of the area under a receiver operating (roc) curve, *Radiology* 143(1): 29–36.

Harris, I, M., Hadden, W., Knowler, W. and Bennet, P. (1987). Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in u.s. population aged 20-74 yr., *Diabetes* 36(4): 523–534.

Hart, W., Goldbaum, M., Cote, B., Kube, P. and Nelson, M. (1997). Automated measurement of retinal vascular tortuosity, In: *Proc. American Medical Informatics Association Annual. Fall Symposium.*, Vol. 63, Nashville, pp. 459–463.

Hauge, S. and Hill, D. (1988). Postural changes in perfusion pressure and retinal arteriolar calibre., *British Journal of Ophthalmology* 72: 253–257.

Henricsson, M., Karlson, C., Ekholm, L., Kaikkonen, P., Sellman, A., Steffert, E. and Tyrberg, M. (2000). Colour slides or digital photography in diabetes screening –a comparison, *Acta Ophthalmologica Scandinavica* 78(2): 164–168.

Hesse, L., Chofflet, J. and Le Mer, Y. (1993). Simulation of the growth pattern of the central retinal artery using a fractal modeling technique, *German Journal of Opthalmology* 2: 116–118.

Hoover, A. and Goldbaum, M. (1998). Fuzzy convergence, In: *Proc. IEEE Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, pp. 716–721.

Hoover, A., Kouznetsova, V. and Goldbaum, M. (1998). Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response, In: *Am. Medical Infomatics Association Conference*, Orlando, pp. 931–935.

Hoover, A., Kouznetzova, V. and Goldbaum, M. (2000). Locating blood vessels in retinal images by piece-wise threshold probing of a matched fileter response, *IEEE Trans. Med. Imag.* 19(3): 203–210.

Ikui, H., Tominaga, Y. and Mimura, K. (1964). The fine structures of the central retinal artery and vein in the optic nerve of the human eye and pathological changes found occurring therein in hypertension and arteriosclerosis, *Acta Soc. Ophthal. Jap.* 68: 899–911.

Iliasova, N., Ustinov, A., Branchevsky, S. and Durasov, A. (1998). Methods for estimating geometric parameters of retinal vessels using diagnostic

images of fundus, In: A. Mikaelian (editor), *Proc SPIE International Soc. Optical Engineering*, SPIE;, Moscow, pp. 316–325.

Ishibashi, T. and Inomata, H. (1993). Ultrastructure of retinal vessels in diabetic patients, *British Journal of Ophthalmology* 77(9): 574–578.

Jasiobedski, P., Taylor, C. and Brunt, J. (1993). Automated analysis of retinal images, *Image and vision computing* 11(3): 139–144.

Javitt, J., Aiello, L., Chiang, Y., Ferris 3rd, F., Canner, J. and Greenfield, S. (1994). Preventive eye care in people with diabetes is cost-saving to the federal government., *Diabetes Care* 17(8): 909–917.

Jiang, X. Y. and Mojon, D. (2003). Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images, *IEEE Trans. Pattern Anal. Mach. Intell.* 25: 131–137.

Kanungo, T., Jaisimha, M., Palmer, J. and Haralick, R. (1993). A quantative methodology for analysing the performance of detection algorithms, In: *Proc. 4th International Conf. on Computer Vision*, Berlin, Germany, May, pp. 247–252.

Katz, N., Goldbaum, M., Nelson, M. and Chaudhuri, S. (1988). An image processing system for automatic retina diagnosis, In: *Proc. SPIE Three-Dimensional Imaging and Remote Sensing Imaging*, Vol. 902, Los Angeles, CA, pp. 131–137.

Kerr, D., Cavan, D., Gold, D., Jennings, B., Davidson, C. and Crick, M. (1997). Beyond retinal screening: digital imaging in the assessment and follow-up of patients with diabetic retinopathy, *Diabetes Nutrition and Metabolism* 11(1): 94.

Kilbride, P., Alexander, K., Fishman, M. and Fishman, G. (1989). Human macular pigment assessed by imaging fundus reflectometry, *Vision Research* 29(6): 663–674.

Kiri, A., Dyer, D., Bressler, N., Bressler, S. and Schachat, P. (1996). Detection of diabetic macular edema: Nidek 3dx stereophotography compared with fundus biomicroscopy, *American Journal of Ophthalmology* 122(5): 654–662.

Klein, R., Klein, B., Magli, Y., Brothers, R., Meuer, S., Moss, S. and Davis, M. (1986). An alternative method of grading diabetic retinopathy, *Ophthalmology* 93(9): 1183–1187.

Klein, R., Klein, B., Moss, S. and Cruickshanks, K. (1995). The wisconsin epidemiologic study of diabetic retinopathy, *Ophthalmology* 102(1): 7–16.

Klein, R., Klein, B., Moss, S., Davis, M. and DeMets, D. (1984). 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* 102(4): 527–532.

Klein, R., Klein, B., Moss, S., Davis, M. and DeMets, D. (1989a). The wisconsin epidemiologic study of diabetic retinopathy: Ix four-year incidence and progression of diabetic retinopathy when age at diagnosis is less than 30 years., *Arch. Ophthalmol.* 107(2): 237–243.

Klein, R., Moss, S., Klein, B., Davis, M. and DeMets, D. (1989b). The wisconsin epidemiologic study of diabetic retinopathy: Xi. the incidence of macular edema, *Ophthalmology* 96(10): 1501–1510.

Kochner, B., Schuhmann, D., Michaelis, M., Mann, G. and Englemeier, K.-H. (1998). Course tracking and contour extraction of retinal vessels

from color fundus photographs: Most efficient use of steerable filters for model based image analysis., In: K. Hanson (editor), *Proc. SPIE Medical Imaging*, SPIE, San Diego, California., pp. 755–761.

Kókai, G., Tóth, Z. and Ványi, R. (1999). Modelling blood vessels of the eye with parametric l–systems using evolutionary algorithms, In: *Joint European Conference on Artificial Intellingence and medical decision making*, Springer Verlag, Aalborg, pp. 433–442.

Kozousek, V., Shen, Z., Gregson, P. and Scott, R. (1992). Automated detection and quantification of venous beading using fourier analysis, *Canadian Journal of Ophthalmology* 27(6): 288–294.

Kristinsson, J., Gottfredsdóttir, M. and Stefansson, E. (1997). Retinal vessel dilation and elongation precedes diabetic macular oedema., *British Journal of Ophthalmology* 81(4): 274–278.

Lalonde, M., Gagnon, L. and Boucher, M. (2000). Non-recursive paired tracking for vessel extraction from retinal images, In: *Proc. Vision Interface 2000*, Montreal, Canada, May, p. na.

Landini, G., Misson, G. and Murray, P. (1993). Fractal analysis of the normal human retinal flourescin angiogram, *Current Eye Research.* 12(1): 23–27.

Landini, G., Murray, P. and Misson, G. (1995). Local connected fractal dimensions and lacunarity analysis of 60 fluorecin angiograms, *Investigative Ophthalmology and Visual Science* 36(13): 2749–55.

Leandro, J., Cesar Jr, R. and Jelinek, H. (2001). Blood vessels segmentation in retina: Preliminary assessment of the mathematical morphology and the wavelet transform techniques, In: *Proc. SIBGRAPI 14th Brazilian Symposium on computer graphics and Image Processing*, Florianópolis-Santa Catalina, Brazil, pp. 84–90.

Lee, E., Rehkopf, P., Warnincki, J., Friberg, T., Finegold, D. and Cape, E. (1997). A new method for assessment of changes in retinal blood flow, *Medical Engineering and Physics* 19(2): 125–130.

Lee, S. (1992). *Identifying retinal vessel networks in ocular fundus images*, Master's thesis, The University of New Mexico.

Lee, S. and Wang, Y. (1999). Automatic retinal image quality assessment and enhancement, In: *SPIE Conference on Image Processing*, Vol. 3661, San Diego, California, pp. 1581–1590.

Lee, S., Wang, Y. and Lee, E. (1999). A computer algorithm for automated detection and quantification of microanuerysms and hemorrhages(hma's) in color retinal images, In: *SPIE Conference on Image Perception and Performance*, Vol. 3663, San Diego, California, pp. 61–71.

Liesenfeld, B., Kohner, E., Piehlmeier, W., Kluthe, S., Porta, M., Bek, T., Obermaier, M., Mayer, H., Mann, G., Holle, R. and Hep, K. (2000). A telemedical approach to the screening of diabetic retinopathy: Digital fundus photography, *Diabetes Care* 23(3): 345–348.

Liu, I. and Sun, Y. (1990). Recursive tracking of vascular trees in angiograms using detection deletion scheme, In: *Proc. IEEE 12th Anual Conf. Eng. in Med. Biol.*, Vol. 12, Philadelphia, USA, November, pp. 0169–0170.

Macskassy, S., Provost, F. and Littman, M. (2003). Confidence bands for roc curves, *CeDER Working Paper*.

Mainster, M. (1990). The fractal properties of retinal vessels: Embryological and clinical implications., *Eye* 4(1): 235–241.

Martínez-Pérez, M., Hughes, A., Stanton, A., Thom, S., Bharath, A. and Parker, K. (1999). Segmentation of retinal blood vessels based on the second directional derivative and region growing, In: *Proc. Internatinal Conf. on Image Processing*, New Jersey, USA, pp. 173–176.

Martínez-Pérez, M., Hughes, A., Stanton, A., Thom, S., Chapman, N., Bharath, A. and Parker, K. (2002). Retinal vascular tree morphology: A semi-automatic quantification, *IEEE Trans. Biomed. Eng.* 49(8): 912–917.

McNitt-Gray, M., Huang, H. and Sayre, J. (1995). Feature selection in the pattern classification problem of digital chest radiograph segmentation., *IEEE Trans. Med. Imag.* 14(3): 537–547.

Meehan, R., Taylor, G., Rock, P., Mader, T., Hunter, N. and Cymerman, A. (1990). An automated method of quantifying retinal vascular response during exposure to novel environmental conditions., *Ophthalmology* 97(7): 875–881.

Mendels, F., Heneghan, C. and Thiran, J.-P. (1999). Identification of the optic disk boundary in retinal images using active contours, In: *Proc. IMVIP (Irish Machine Vision and Image Processing) Conference 1999*, Dublin, Eire, September, pp. 103–115.

Michaelson, I. (1948). The mode of development of the vascular system of the retina, with some observations on its significance for certain retinal diseases., *Trans Opthalmic Soceity United Kingdom* 68: 137–180.

National Institute Clinical Excelence (2002). Management of type 2 diabetes retinopathy - early management and screening (guideline e), *Technical Report N0058 70k Feb 02 (ABA)*, National Institute Clinical Excelence. ISBN: 1-84257-146-X.

National Screening Committee (2000). *Second Report of the United Kingdom National Screening Committee*, Department of Health, Po. Box 777, London SE1 6XH, England, United Kindom.

Newsom, R., Sullivan, P., Rassam, S., Jagoe, R. and Kohner, E. (1992). Retinal vessel measurement: comparison between observer and computer driven methods, *Graefe's Arch Clin Exp Ophthalmology* 230(4): 221–225.

Okazaki, K. and Tamura, S. (1983). Spherical sharding correction of eye fundus image, In: *Proc. of International Conference on Systems, Man and Cybernetics*, Bombay, India, pp. 1084–1087.

Osareh, A., Mirmehdi, M., Thomas, B. and Markham, R. (2002). Classification and localisation of diabetic-related eye disease., In: A. Heyden, G. Sparr, M. Nielsen and P. Johansen (editors), *7th European Conference on Computer Vision*, Copenhagen, Denmark, May, pp. 502–516.

Page, M., Howard, R., O'Brien, J., Buxton-Thomas, M. and Pickering, A. (1996). Use of neural networks in brain spect to diagnose alzheimer's disease., *Journal of Nuclear Medicine* 37(2): 195–200.

Panico, J. and Sterling, P. (1995). Retinal neurons and vessels are not fractal but space-filling, *The Journal of Comparative Neurology* 361(3): 479–490.

Patel, V. (1995). *Diabetic Retinopathy: Haemodynamic and clinical Factors in the Pathogenesis.*, Verlag: Josef Eul, Koln. ISBN: 3-89012-432-1.

Peli, E. (1993). Enhancement of retinal images: Pros and problems, *Neuroscience and Biobehavioral Reviews* 17(4): 477–482.

Poli, R. and Valli, G. (1997). An algorithm for real-time vessel enhancement and detection, *Computer Methods and Programs in Biomedicine* 52(1): 1–22.

Prakash, M. and Narashimah Murty, N. (1995). A genetic approach for selection of (near-) optimal subsets of principal components for discrimination, *Pattern Recognition Letters* 16(8): 781–787.

Rambhia, A. (1999). *Automatic extraction of vascular tree from retinal flourecin angiograms*, Master's thesis, University of Washington.

Rassam, S., Patel, V., Brinchmann-Hansen, O., Engvold, O. and Kohner, E. (1994). Accurate vessel width measurement from fundus photographs: a new concept., *British Journal of Ophthalmology* 78(1): 24–29.

Robbins, A., Hurley, L., Dudenhoefer, E. and Chao, S. (2001). Performance characteristics of digital fundus photography as a screening test for diabetic retinopathy in a low-risk population, *Diabetes Technology and Therapeutics* 3(2): 193–199.

Ruwaard, D., Gijsen, R., Bartelds, A. I. M., Hirasing, R. A., Verkleij, H. and Kromhout, D. (1996). Is the incidence of diabetes increasing in all age-groups in the netherlands?: Results of the second study in the dutch sentinel practice network, *Diabetes Care* 19(3): 214–218.

Sahiner, B., Chan, H.-P., Wei, D., Petrick, N., Helvie, M., Adler, D. and Goodsitt, M. (1996). Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue, *Med. Phys.* 23(10): 1671–1684.

Saude, T. (1993). *Ocular Anatomy and Physiology*, Blackwell Scientific Publications, Oxford.

Schalkoff, R. (1989). *Digital Image Processing And Computer Vision*, Wiley and Sons, Inc., New York.

Scott, J. and Palmer, E. (1993). Do diagnostic algorithms always produce a uniform lung scan interpretation ?, *Journal of Nuclear Medicine* 34(4): 661–665.

Scott, J., Fisher, R. and Palmer, E. (1996). Neural networks in ventilation-perfusion imaging; part ii. effects of interpretive variability, *Radilogy* 198(3): 707–713.

Shen, Z., Gregson, P., Cheng, H. and Kozousek, V. (1991). Automated grading of venous beading : An algorithm and parallel implementation, In: K. Tzou and T. Koga (editors), *Proc. SPIE Visual Communications and Image Processing 91*, Boston, MA, pp. 632–640.

Shiba, T., Yamamoto, T., Seki, U., Usugi, N., Fujita, K., Sato, Y., Terada, H., Sekihara, H. and Hagura, R. (2002). Screening and follow-up of diabetic retinopathy using a new mosaic 9-field fundus photography system, *Diabetes Research and Clinical Practice* 55(1): 49–59.

Shin, D., Kaiser, R., Lee, M. and Berger, J. (1999). Fundus image chage analysis : Geometric and radiometric normalisation, In: *Proc. SPIE Conference on Ophthalmic Technologies*, Vol. 3591, San Jose, CA, pp. 129–136.

Sinthanayothin, C. (1999). *Image analysis for automatic diagnosis of diabetic retinopathy*, PhD thesis, Department of Physics, Kings College London.

Sinthanayothin, C., Boyce, J., Cook, H. and Williamson, T. (1999). Automated localisation of the optic disk, fovea, and retinal blood vessels from digital colour fundus images., *British Journal Ophthalmology* 83(8): 902–910.

Spencer, T. (1992). *Digital Imaging of the Retina*, PhD thesis, Department of Bioengineering, Aberdeen University.

Staal, J., Kalitzin, S. N., van Ginneken, B., Abramoff, M. and Viergever, M. (2002). Classifying convex sets for vessel detection in retinal images, In: *International Symposium on Biomedical Imaging*, Washington D.C., July, pp. 269–272.

Stitt, A., Gardiner, T. and Archer, D. (1995). Histological and ultrastructural investigation of retinal microaneurysm development in diabetic patients., *British Journal of Ophthalmology* 79(4): 362–367.

Suzuki, Y. (1995). Direct measurement of retinal vessel diameter - comparison with microdesiometric methods based on fundus photographs, *Survey of Ophthalmology* 39(S1): 57–65.

Tamura, S., Okamoto, Y. and Yanashima, K. (1988). Zero-crossing interval correction in tracing eye-fundus blood vessels, *Pattern Recognition* 21(3): 227–233.

Tascini, G., Passerini, G., Puliti, P. and Zingaretti, P. (1993). Robust approach to occular fundus image analysis, In: *Proc. SPIE Biomed Image Process. and Biomed. Visualization*, Vol. 1905, San Jose, CA, USA., pp. 638–647.

Tolias, Y. and Panas, S. (1998). A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering, *IEEE Trans. Med. Imag.* 17(2): 263–273.

Toussaint, D. and Danis, P. (1970). Vascular density coefficeints in normal human retina and relationship to distance to disk, *Arch Opthal* 83(March): 281–285.

Walter, T. and Kllein, J.-C. (2001). Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques., In: *Second International Symposium on Medical Data Analysis*, Madrid, Spain, May, pp. 282–287.

Wang, Y. and S.C., L. (1997). A fast method for automated detection of blood vessels in retinal images, In: *Proc. Asilomar Conf. on Circuits Systems and Computing*, Pacific Groove, California, USA, November, pp. 1700–1704.

Wang, Y., Toonen, H. and Meyer-Ebrecht, D. (1990). A new method of automatic tracking, measuring blood vessels in retinal image., In: *Ann. International Conf. of the IEEE Eng. Medicine and Biology Society*, Vol. 12, Philadelphia, Pen. USA, pp. 0174–0175.

West, G., Brown, J. and Enquist, B. (1997). A general model for the origin of allometric scalling laws in biology, *Science* 276(5309): 122–126.

WHO (1989). Diabetes care and research in europe: The saint vincent declaration., *Technical Report ICP/CLR 034 6884s*, World Health Organisation, Regional Office for Europe; and The International Diabetes Federation (IDF).

Wilson, C., Steffánsson, E., Klombers, L., Hubbard, L., Kaufman, S. and Ferris 3rd, F.L. (1988). Optic disk neovascularisation and retinal vessel diameter in diabetic retinopathy, *American Journal of Ophthalmology* pp. 131–134.

Wu, Y., Giger, M.L. nd Doi, K., Vyborny, C., Schmidth, R. and Metz, C. (1993). Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer, *Radiology* 187(1): 81–87.

Yamamoto, S. and Yokouchi, H. (1976). Automatic recognition of color fundus photographs, In: *Digital Processing of Biomedical images*, Plenum Press, New York, pp. 385–398. ISBN 0-306-309687-X.

Yang, C.-W., Ma, D.-J., Wang, C.-M., Wen, C.-H., Lo, C.-S. and Chang, C.-I. (2000). Computer-aided diagnostic detection system of venous beading in retinal images, *Optical Engineering* 39(5): 1293–1303.

Ye, D. and Zheng, L. (1995). Fundus image processing and feature classification based on mathematical morphology method., In: *Proc. of the Canadian Medical and Biological Engineers Conference*, Vol. 2, Montreal, pp. 1015–1016.

Young, S., George, L., Lusty, J. and Owens, D. (1997). A new screening tool for diabetic retinopathy: the canon cr5 45nm retinal camera with frost medical software ris-lite digital imaging system., *Journal of Audiovisual Media in Medicine* 20(1): 11–14.

Yu, J., Hung, B. and Sun, H. (1990). Automatic recognition of retinopathy from retinal images, In: *Proc. of the Twelfth Annual International Conference on the IEEE Engineering and Medicine and Biology Society*, Philadelphia.

Yulong, M. and Dingru, X. (1990). Recognizing the glaucoma from ocular fundus image by image processing, In: *Proc. of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 12, Philadelphia, pp. 178–179.

Zahlmann, G., Kochner, B., Ugi, I., Schulhmann, D., Liesenfeld, B., Wegner, A., Obermaier, M. and Mertz, M. (2000). Hybrid fuzzy image processing for situation assessment: A knowledge–based sysetm for early detection of diabetic retinopathy, *IEEE Engineering in medicine and biology* 19(1): 76–83.

Zana, F. and Klien, J. (2001). Segmentation of vessel-like patterns using mathematical morphology and curvature elevation, *IEEE Trans. Img. Processing* 10(7): 1010–19.

Zhou, L., Rzeszotarski, M., Singerman, L. and Chokreff, J. (1994). The detection and quantification of retinopathy using digital angiograms, *IEEE Trans. Med. Imag.* 13(4): 619–26.

Zou, K., Hall, W. and Shapiro, D. (1997). Smooth non-parametric receiver operating characteristic (roc) curves for continuous diagnostic tests, *Statistics In Medicine* 16(19): 2143–2156.