Diabetic retinal imaging: methods in automatic processing

A thesis submitted to The University of Manchester for the degree of MPhil in the Faculty of Life Sciences

YEAR OF SUBMISSION 2014

CANDIDATE’S NAME Greg Russell
**Abstract**

Fundus image based screening for diabetic retinopathy is offered to all diabetic patients aged 12 and older. This has proven to be an effective procedure for the early detection/diagnosis of diabetic retinopathy and forms the basis of current treatment plans. However, the increasing number of diabetic patients is putting a strain on the NHS. Computer based tools to aid detection of/grade diabetic pathologies are currently under development. In this MPhil a novel database of fundus is described. Many of whom also possess comorbidities such as glaucoma or hypertension. Retinal vessel masks were extracted by hand to establish accurate high-resolution images to test automatic vessel extraction algorithms on. Two previously published automatic vessel segmentation algorithms were tested on this database.

This collection of images accurately represents the variety of fundus images a retinal grader can expect to encounter in a regional screening program in the UK. In addition retinal image quality can be significantly degraded by media opacities, limiting the diagnostic potential of retinal images. The amount of scattering increases with age and with some pathologies (e.g. cataract). Even though a large body of work exists on the enhancement of images recorded in poor visibility very little has been done on reducing the degradation of retinal images by intraocular scattering. In this thesis a defogging filter designed to enhance image clarity was applied to fundus images that had previously been graded as inadequate. 12% of these images were found to be assessable after filtering suggesting that a ‘cataract filter’ of this type may be beneficial in diabetic retinopathy screening programmes.

“The only thing worse than being blind is having sight but no vision.”

Helen Keller
Acknowledgements

From working in an industry dedicated to providing services of retinal screening for 14 years, to becoming a part time MPhil student for the first time after many years of education at one of the most prestigious University’s in England can certainly be a challenge and quite daunting. I would therefore like to thank my supervisors Dr Vincent Nourrit and for firstly giving me an opportunity to undertake this work and Dr Niall McLoughlin who offered great support when Vincent took up an excellent opportunity to work in France, and could no longer be my supervisor (although has still offered support from overseas). Niall has given me comprehensive support all though my days as a part time student at Manchester. Also thanks to Dr John Oakley from the faculty of Engineering and Physical Sciences at the University of Manchester for all his support on this project and advice, and all of the students and staff at faculty of eye and vision sciences who have inspired me in this project. Finally I would like to offer special thanks to my boss Phil Kirby to have the time to complete my MPhil and my colleague Mr Adam Ellis who assisted with software integration of the Cleavue filter into our grading software. This then allowed me to alter parameters needed for my tests and batch processing of images.
Contents

Chapter 1: Introduction ........................................................................................................... 8
  1.1 Rise in the number of patients being screened ............................................................ 9
  1.2 Non-mydriatic Diabetic Retinopathy Screening .......................................................... 11
  1.3 Diabetic retinopathy .................................................................................................... 12
    1.3.1 R0 Classification .................................................................................................. 13
    1.3.2 R1 Classification .................................................................................................. 13
    1.3.3 R2 Classification .................................................................................................. 14
    1.3.4 R3A Classification ............................................................................................... 15
    1.3.5 R3S Classification ............................................................................................... 16
    1.3.6 M1 Classification ............................................................................................... 17
    1.3.7 Other conditions ............................................................................................... 18
  1.4 Benefits of early diagnosis/detection ............................................................................ 18
  1.5 Recent advances in diabetic retinal screening ............................................................ 24
  1.6 Photography ................................................................................................................ 26
  1.7 Grading Pathway & Outcomes ..................................................................................... 28
  1.8 Portable hand held retinal cameras ............................................................................. 30
  1.9 Online Databases for testing algorithms for automated screening purposes .......... 31
    1.9.1 DRIVE database ................................................................................................. 31
    1.9.2 STARE database ............................................................................................... 32
    1.9.3 ARIA online database ....................................................................................... 33
    1.9.4 VICAVR database ............................................................................................. 34
    1.9.5 Messidor database ............................................................................................. 34
    1.9.6 ImageRet database ............................................................................................ 36
    1.9.7 REVIEW image set ............................................................................................ 36
    1.9.8 ROC database ................................................................................................... 37
    1.9.9 The need for a realistic database ......................................................................... 38
  1.10 Current limitations with digital fundus images ............................................................ 39
  1.11 Gauging assessability of retinal images ...................................................................... 46
  1.12 Possible solution: The cataract filter algorithm (Clearvue) ......................................... 50
  1.13 Overview of the rest of the thesis ............................................................................... 52

Chapter 2: DR HAGIS – A new realistic DR Database for the Automatic
Extraction of Retinal Surface Vessels .................................................................................... 54

  2.1 The need for a realistic DR database ........................................................................... 54
  2.2 DR HAGIS database .................................................................................................. 54
  2.3 Ground truth images .................................................................................................. 55
  2.4 Initial retinal vessel segmentation results .................................................................... 58
  2.5 Discussion .................................................................................................................. 62

Chapter 3: Filtering the effects of cataracts ................................................................. 63

  3.1 Background ............................................................................................................... 63
  3.2 Current image enhancement techniques to improve image quality ......................... 64
  3.3 Methods ...................................................................................................................... 65
3.4 A pilot test of the ‘cataract filter’ algorithm ........................................ 67
3.5 Results .................................................................................................... 69
3.6 Discussion .............................................................................................. 72
Chapter 4: Conclusions and future work .................................................. 74

4.1 DR HAGIS database ............................................................................ 75
4.2 Issues with the preliminary cataract filter study ................................. 75
4.3 Further cataract filter studies ............................................................... 76
4.4 Conclusions ........................................................................................... 78

References .................................................................................................. 80
Appendices ...................................................................................................... 87
Table of figures

Figure 1 An example image of a patient who would be classed as R0 .................. 13
Figure 2 An example of an R1 classification .................................................. 14
Figure 3 An example of the R2 classification .................................................. 14
Figure 4 An example of the R3A classification .............................................. 15
Figure 5 An example of stable R3S DR ......................................................... 16
Figure 6 An example of a retina that would be graded M1 .............................. 17
Figure 7 Image showing a central retinal vein occlusion .............................. 18
Figure 8 Four images mandated by the ENSPDR ........................................ 26
Figure 9 Health Intelligence measurement tool ............................................. 27
Figure 10 The Common pathway for grading DR fundus images .................... 29
Figure 11 Optomed Smart Scope Pro portable retinal camera ........................ 31
Figure 12 Two example images from the DRIVE databases ............................ 32
Figure 13 Two example 35 degree images from the STARE database .......... 33
Figure 14 Two control images taken from the ARIA online databases .......... 33
Figure 15 An example from the VICAVR database ....................................... 34
Figure 16 Images taken from the Messidor database ..................................... 35
Figure 17 Images taken from the ImageRet database ...................................... 35
Figure 18 An image from the REVIEW database ........................................... 36
Figure 19 An example retinal image from the ROC database ......................... 37
Figure 20 An example showing differences of white balance, exposure and barrel distortion .......................................................... 38
Figure 21 The spatial resolution of film in units equivalent to digital camera megapixels .......................................................... 41
Figure 22 Estimated costs of clinical malpractice resulting from misdiagnosis .... 42
Figure 23 Effects of image manipulation ...................................................... 43
Figure 24 Image showing colour distortions ................................................. 44
Figure 25 Retinal images showing a microaneurysm ..................................... 45
Figure 26 The progression of cataract in a patient over a 5 year period ........... 46
Figure 27 Example images from the Scottish grading scheme ....................... 48
Figure 28 Screening outcomes for patients in the East Anglia Eye Screening Programme ........................................... 49
Figure 29 Images depicting the similarities between pictures taken on a foggy day and fundus images taken through a media opacity .................................. 50
Figure 30 An airborne image degraded by haze ............................................ 51
Figure 31 An example of segmentation method one ...................................... 56
Figure 32 An example of segmentation method two ..................................... 57
Figure 33 An example of the final image tracing technique ............................ 57
Figure 34 Example fundus images from the Dr Hagis database with automatic segmentation using the intensity based approach .......................... 59
Figure 35 Example fundus images from the Dr Hagis database with automatic segmentation using a two scale gabor filter based approach .................. 60
Figure 36 An example of an unassessable retinal image ................................ 63
Figure 37 Example of simulated clear and foggy conditions ........................ 66
Figure 38 The five images selected for detailed investigation ......................... 70
Figure 39 An example of how varying the parameters of the cataract filter affect the quality of a single unassessable image ......................................... 72
Figure 40 Comparison of unassessable and the ‘Clearvue’ filtered image .......... 73
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DD,2DD</td>
<td>One, two optic Disc Diameters</td>
</tr>
<tr>
<td>A/V ratio</td>
<td>Artery to Vein ratio</td>
</tr>
<tr>
<td>AAO</td>
<td>American Academy of Ophthalmology</td>
</tr>
<tr>
<td>AMD</td>
<td>Age-related Macular Degeneration</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge Coupled Device</td>
</tr>
<tr>
<td>CE</td>
<td>European Certificate</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>CRVO</td>
<td>Central Retinal Vein occlusion</td>
</tr>
<tr>
<td>CWS</td>
<td>Cotton Wool Spots</td>
</tr>
<tr>
<td>DM</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>DNA</td>
<td>Did Not Attend</td>
</tr>
<tr>
<td>DR</td>
<td>Diabetic Retinopathy</td>
</tr>
<tr>
<td>DSLR</td>
<td>Digital Single Lens Reflex</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated Glomerular Filtration Rate</td>
</tr>
<tr>
<td>FDA</td>
<td>Federal Drugs Administration</td>
</tr>
<tr>
<td>FA</td>
<td>Fundus Fluorescein Angiography</td>
</tr>
<tr>
<td>FFDS,SFDG</td>
<td>First, Second Full Disease Grader</td>
</tr>
<tr>
<td>GBP</td>
<td>Great British Pound</td>
</tr>
<tr>
<td>GIMP</td>
<td>Gnu Image Manipulation Program</td>
</tr>
<tr>
<td>HES</td>
<td>Hospital Eye Service</td>
</tr>
<tr>
<td>ICER</td>
<td>Incremental Cost Effectiveness Ratio</td>
</tr>
<tr>
<td>ICO</td>
<td>International Council for Ophthalmology ICO</td>
</tr>
<tr>
<td>IRMA</td>
<td>Intra-retinal Microvascular Abnormalities</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standards Organisation</td>
</tr>
<tr>
<td>LogMAR</td>
<td>Logarithm of the Minimum Angle of Resolution</td>
</tr>
<tr>
<td>M0,1</td>
<td>Maculopathy Grade</td>
</tr>
<tr>
<td>MA</td>
<td>Microaneurysms</td>
</tr>
<tr>
<td>MDT</td>
<td>Multi-Disciplinary Team Meeting</td>
</tr>
<tr>
<td>NHS DESP</td>
<td>NHS Diabetic Eye Screening Programme</td>
</tr>
<tr>
<td>NSC</td>
<td>National Screening Committee</td>
</tr>
<tr>
<td>NSF</td>
<td>National Service Framework</td>
</tr>
<tr>
<td>NVD</td>
<td>New Vessel growth at the Disc (NVD)</td>
</tr>
<tr>
<td>NVE</td>
<td>New Vessel growth Elsewhere in the retina</td>
</tr>
<tr>
<td>P0,P1</td>
<td>Photocoagulation grade</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality Adjusted Life Year</td>
</tr>
<tr>
<td>R0,1,2,3A,3S</td>
<td>Retinopathy Grade</td>
</tr>
<tr>
<td>ROG</td>
<td>Referral Outcome Grade (er)</td>
</tr>
<tr>
<td>ROP</td>
<td>Retinopathy Of Prematurity</td>
</tr>
<tr>
<td>RPE</td>
<td>Retinal Pigmented Epithelium</td>
</tr>
<tr>
<td>SLBM</td>
<td>Slit Lamp Bio-microscopy</td>
</tr>
<tr>
<td>SLR</td>
<td>Single Lens Reflex</td>
</tr>
<tr>
<td>STDR</td>
<td>Sight Threatening Diabetic Retinopathy</td>
</tr>
<tr>
<td>UAE</td>
<td>United Arab Emirates</td>
</tr>
<tr>
<td>UKPDS</td>
<td>United Kingdom Prospective Diabetes Study</td>
</tr>
<tr>
<td>USD</td>
<td>United States dollars</td>
</tr>
<tr>
<td>VA</td>
<td>Visual Acuity</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular Endothelial Growth Factor</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

The following four notes on copyright and the ownership of intellectual property rights must be included as written below:

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses

v. The work carried out in this thesis is the totally original work of the author, except where acknowledged by references.
Chapter 1: Introduction

Diabetic retinopathy is damage to the retina caused by complications of diabetes. It can eventually lead to blindness and affects up to 80 percent of all patients who have had diabetes for 10 years or more (Kertes and Johnson, 2007). Diabetic retinopathy is currently the only ocular disease that is actively screened for in the UK with annual screening being offered to all diabetics over the age of 12. However the global increase in the number of patients diagnosed with diabetes mellitus coupled with the increasing costs and demands on this screening system means methods to semi-automate or aid screeners are much in demand. DR is no longer the leading cause of blindness in the working age population in England and this can be attributed to the well-established English diabetic eye screening programme (Liew et al, 2014).

This thesis will present the results of two preliminary studies: one outlining a new database of fundus images taken from diabetic patients being screened under the NHS Diabetic Eye Screening Programme (DESP) that can be used as a test-bed for automatic or semi-automatic vessel extraction algorithms to aid in clinical analysis. A second study investigates the potential of a new software filter that may help in cleaning-up low quality, so called inadequate, screening images so that fewer patients will need to wait for results or be referred on for more costly investigations.

Before detailing these studies the need for developments in retinal screening and how current screening programs operate are presented. Thankfully a number of effective treatments exist for diabetic retinopathy if it is detected early enough. These will be briefly reviewed along with recent advances in diabetic retinal screening. Finally in this chapter the concept of image adequacy/assessibility will be reviewed.
along with some of the factors that control this before a new software filter which demonstrates some promise in helping to clean-up at least some of the poorer quality fundus images is presented.

1.1 Rise in the number of patients being screened

The dramatic ageing of the world population coupled with the massive increase in the incidence of diabetes worldwide is increasing the strain on health services to provide adequate retinal screening for diabetic retinopathy, amongst other age-related ocular diseases (Zimmet et al., 2001; Voleti and Hubschman, 2013).

Diabetes mellitus (diabetes) is a global epidemic. In 2011, there were 366 million people (8.3% of the world’s adult population) living with diabetes; by 2030 this will have increased by 54% to 552 million (International Diabetes Federation, 2011). More than 80% of deaths associated with diabetes occur in low- and middle-income countries. The World Health Organisation (WHO) projects that diabetes will be the 7th leading cause of death by 2030. Diabetes increases the risk of a range of eye diseases, but the main cause of blindness associated with diabetes is diabetic retinopathy (DR). While DR is not currently the primary cause of avoidable blindness, it has the capacity to become the leading cause of blindness in the next 20 years and it will affect the poorest people the most - already 80% of people with diabetes live in low-, middle-income countries (World Health Organization, 2014). Assuming an annual screening model is used then based on a 54% increase in the global diabetic population by 2030 almost 35 exams per second will be needed every second of every day in order to screen the diabetic population of the world. This prediction is compounded by the fact that there will be less than a 2% growth in the number of ophthalmologists by 2030 (Iapb.org, 2015).
The limited availability of a trained workforce limits service quality and reach globally. An estimation of the worldwide economic and health burden of visual impairment was found to be in the region of approximately 3 trillion USD (Gordois et al, 2011). Currently there is a 3 billion USD spending devoted to screening for blinding eye diseases according to a report from AMD Alliance International.

With the global epidemic of diabetes (Whiting et al, 2011), the worldwide implementation of eye screening programmes becomes more important. At the same time, the cost of establishing and running such screening programmes is considerable and rising. Annual screening for Diabetic Retinopathy (DR) began in Iceland in 1980. In 1994 they proposed biennial screening of patients without retinopathy and annual screening for patients with signs of DR. This type of biennial screening has proven to be safe and effective, while reducing costs (Kristinsson et al, 1995; Olafsdottir and Stefansson, 2007).

In countries/regions with no provision for DR screening there is the obvious risk to diabetic patients in developing DR but also from other sight threatening conditions. One private UK based screening provider for the NHS called Health Intelligence screens DR images for eight different conditions needing urgent referral, twelve different conditions that need GP referral, and seven different conditions that need notification for the patient records. DR screening has a massive impact on saving people from loss of vision.

In countries/regions with DR screening, existing arrangements such as the inappropriately high screening frequency for low risk patients can also lead to increased disease burden in other patients. The minority of patients who are at high
risk of developing Sight Threatening Retinopathy (STR) may presently experience delayed provision of the correct treatment due to inappropriate screening.

While the costs per patient of DR screening are rather low the prevalence of diabetes is rising and with it the total cost of DR screening as a whole. This makes it important to make the screening programme as cost effective and efficient as possible. The attainable cost savings of personalized DR screening vary across countries with population size. In the UK the cost for providing a full screening service is on average around £31 GBP based on a survey of 4 separate services charging £15, £17, £45, £50 GBP per patient per annum (depending on local population size and screening model used e.g. optometrist, technicians subcontracted for delivery of the NHS service). In addition we can assume that any personalized approach is likely to also promote compliance with treatment, adding to efficiency gains in health care.

1.2 Non-mydriatic Diabetic Retinopathy Screening

A photographic system has advantages over individual specialist retinal assessment by an ophthalmologist /optometrist. First, technicians can perform camera screening. Second, a camera can be flown to an isolated community. Third, images can be archived to permit comparison over time. There have been a number of proposed screening methods for DR based on digital retinal photography, the obvious advantage with digital imaging is the reduced cost, good expectations of clarity of image and ability to view images remotely (Kim et al 2007; Boucher et al 2003; Sharp et al 2003). Two-field fundus photography is the most common method deployed within the UK. In England the national screening programme recommends that two images are taken of each eye (Lacey and Taylor, 2014). In Scotland a single image is taken and patients dilated only when required despite the fact that this has been shown
to be less effective at detection of clinical features (Aldington et al., 1995, Ding et al., 2012, Murgatroyd et al, 2004).

In many countries, physicians or optometrists conduct DR screening during their clinical practice. But they may be limited by a lack of dedicated time and these professionals are theoretically more costly. In the UK, non-physician graders and screeners have proven to be able to provide good detection of DR and maculopathy from fundus photographs and are both clinically and cost-effective.

In the UK screeners are trained on how to operate the fundus camera and take photographs. The patient’s pupils will be dilated only if they are too small for adequate imaging and after consent is obtained from the patients. Two-field retinal photography with one image centred on the macula and encompassing the optic disc and a second image containing a nasal view centred on the optic disc and encompassing a second view of the macular is obtained from both eyes of each participant. This two-field approach supports the correct identification of maculopathy as fine exudates and haemorrhages which may be difficult to see using other techniques such as Optical Coherence Tomography (OCT). The photographs are stored digitally and copies of the grading report are kept electronically along with the patient’s screening data.

1.3 Diabetic retinopathy

Diabetic retinopathy causes damage to blood vessels of the eyes; diabetic retinopathy does not usually affect sight until these changes are advanced. At least 75% of people who have diabetes for more than 20 years will have some form of DR (World Health Organization, 2014). Annual screening is an effective way of preventing sight loss
caused by diabetes, but if no treatment is given this can result in vision loss. Known risk factors that increase the risk of developing DR include duration and type of diabetes, poorly controlled HbA1c, high blood pressure, retinopathy stage and gender. DR develops gradually and the severity of DR is classified using grading classifications that correspond to disease progression.

1.3.1 R0 classification

The R0 classification occurs when no diabetic retinopathy or maculopathy is present, and there is little chance that the patients vision is effected other than possible blurring from lens swelling due to hyperglycemic episodes or cataract formation due to poor sorbital metabolism (see Figure 1 for an example image of a patient with a R0 classification).

![Figure 1: An example image of a patient who would be classed as R0. There are no visible signs of DR. The image is clearly visualised as we can see the retinal vessels within 1 disk diameter of the centre of both disc and fovea.](image)

1.3.2 R1 classification

The next stage in disease progression is a R1 classification. This is not too serious, however there is some noticeable pericytic loss in the blood vessels, perhaps a few
small microaneurisms (circled in yellow in Figure 2) and/or superficial flame haemorrhages may be visible in the retina. There may also be some exudates where leakage takes place. The patent is notified about these changes but no further action will be taken until next year’s screen. Again, as with the R0 classification, there is little chance of the patients vision being affected over all (see Figure 2 for an example of the fundus image of a patient with an R1 classification).

Figure 2: An example of an R1 classification. Such an image may contain background microaneurysm(s) retinal haemorrhage(s) ± any exudate not within the definition of maculopathy. This image shows a yellow circle around a microaneurysm in an otherwise ‘quiet’ retina with no other observable signs of retinopathy.

1.3.3 R2 Classification

R2 Classification is defined by substantial amounts of pericytic damage that lead to deeper rounder haemorrhages. Axoplasmic flow blockage may form Cotton Wool Spots (CWS) that result in ischemic areas. Typically this will stimulate a release of vascular endothelial growth factor hormone (VEGF) in the retina leading to intra retinal microvascular abnormalities (IRMA). These are tiny new vessels trapped within the retina (circled in yellow in Figure 3), venous beading and venous loops.
Multiple deep, round or blot haemorrhages are also symptomatic of R2. Patients are refereed for non-urgent referral to the hospital eye department as there is significant risk to their vision. Patient may already be suffering with some vision loss.

Figure 3: An example of the R2 classification. This figures depicts Pre-proliferative venous beading, venous loops or reduplication, intra-retinal microvascular abnormalities (IRMA), multiple deep, round or blot haemorrhages, and some cotton wool spots seen inside the yellow circles.

Figure 4: This image demonstrates the R3A classification. Proliferative new vessels on disc (NVD) new vessels elsewhere (NVE) and pre-retinal haemorrhage (circled in yellow on the right).
1.3.4 R3A Classification

A fundus image that is classed as R3A typically contains new vessels at the disc (NVD) or elsewhere (NVE) triggered by VEGF. These new vessels can proliferate on the surface of the retina growing towards ischemic areas. This deranged repairing mechanism is the cause of vitreous and sub-haloid haemorrhages. New vessels are often supported by a structure of fibrous tissue, these can adhere to parts of the retina causing significant traction, detaching or distorting of the retina. Patients can become blind if haemorrhage occurs as a result (circled in yellow in Figure 4).

![Fundus Image](image)

**Figure 5:** An example of stable R3S DR. Proliferative features including fibrous tissue (with no traction) and round laser photocoagulation burns that indicate previous laser treatment are present. There are no bleeds or other retinal features that have worsened since discharge from the eye clinic.

1.3.5 R3S Classification

The R3S classification is made if the proliferate features appear stable in comparison to the previous year’s images (exhibiting no change) with the inclusion of a photocoagulation grade (P1) (see Figure 5). These patients can stay in surveillance indefinitely until such time as the referral outcome grader (ROG) adds them back into screening. Doing so will prevent the patients from being unnecessarily referred.
back to the hospital eye service by inexperienced graders unless reactivation of stable treated retinopathy or referable non diabetic retinopathy is found.

1.3.6 M1 Classification

Finally the M1 Classification indicates maculopathy, pathology associated with the macula. Maculopathy, is typically as a result of ischemia (no blood circulation) or oedema (water logging from blood plasma) affecting the macula. Ischemic maculopathy is impossible to treat because the starved blood cells die. Oedema on the other hand may respond to treatment. Microaneurysms and haemorrhages overlaying the macular do not necessarily mean significant leakage is occurring, however if these features are within one disc diameter of the centre of the fovea with an accompanying visual acuity (VA) of worse than 6/12 there may be a relationship. This patient would be referred on for urgent laser treatment.

Figure 6: Shows an example of a retina that would be graded M1. Exudates (lipids) can be seen within 1 disc diameter (DD) of the centre of the fovea. These features typically appear as a line towards the centre of the fovea (tracking), or as a ring (circinate) around the centre of vessel plasma leakage.
1.3.7 Other conditions

On occasion, as a result of patients attending their diabetic eye-screening test, other non-diabetic eye conditions are detected. Some of the conditions may require a referral to the hospital eye service such as the image of a central retinal vein occlusion (CRVO) (see Figure 7). Others will be directed to their GP practice, as they may already be aware of their condition. Although these features may have nothing to do with diabetic retinopathy, the patient will be referred for a hospital eye service appointment if this is required.

![Image of central retinal vein occlusion](image)

Figure 7: Image showing a central retinal vein occlusion. Within this image we can see the optic disc, and many superficial flame shaped haemorrhages (typical of hypertensive bleeds) radiating from the optic disc out along the nerve fibres giving the illusion of a flame.

1.4 Benefits of early diagnosis/detection

People with DR whose sight is at risk can be treated, typically with laser or intra-vitreal drug injections (Meads and Hyde, 2003). Because the effects of diabetes on the eye are generally asymptomatic, many people with diabetes are not aware of any
changes. If left unmanaged DR can drastically affect their vision possibly leading to blindness. The vast majority of patients in countries will poorly established screening programmes that develop DR have no symptoms until the late stages when it affects their central vision, by which time it may be too late for effective preventive treatment.

DR screening is designed to detect early (asymptomatic) sight threatening changes to optimize the timing of preventive treatment. A reduction by one-third or more in new blindness due to diabetes was adopted as one of the key five-year targets in the St Vincent declaration in 1990 (Diabetes Care and Research in Europe: The Saint Vincent Declaration, 1990) which declares that “annual eye examinations are recommended for patients with diabetes (and every other year for persons with excellent glycemic control and no retinopathy at the previous examination in certain contexts)”.

Members of this consortium have led the transition to biennial screening for diabetic patients without retinopathy and this is now recommended by the International Council for Ophthalmology (Chakrabarti et al, 2012). However the DR screening programme in the UK still screens all diabetic patients annually.

Regular screening of diabetic patients for sight threatening retinopathy (STR) could optimise the timing of laser treatment. Making sure that patients are treated at an optimal stage of the disease can improve the sight preservation to about 75%. In Iceland diabetic blindness prevalence decreased 4-5 fold after the advent of systematic diabetic eye screening and similar success has been seen in Denmark Diabetes care (Stefánsson et al, 2000).
Saddine et al (2008) forecast that the number of people with DR in the US will increase 3 fold from 2005 to 2020 further illustrating the importance for establishing a diabetic retinal screening programme in the US. In the United Arab Emirates (UAE) the true potential burden of visual impairment is only just being realised. In the US nearly 8.1 million people have undiagnosed diabetes (American Diabetes Association, 2014). In the UK it is thought to be 1 in 70 (Diabetes.org.uk, 2014). In the UAE there are 745,940 diabetics (International Diabetes Federation, 2014) and it is estimated in approximately 304,000 of these cases the condition has not yet been diagnosed. There are in addition an estimated 934,300 people with impaired glucose tolerance, a pre-diabetic state of hyperglycaemia, or elevated levels of blood sugar (International Diabetes Federation, 2014).

In comparison to the UAE the healthcare system in the US is becoming more inclusive of screening for DR. Approximately 1 in 29 Americans 40 years and older has DR of any severity, and 1 in 132 people have vision threatening DR (Congdon N, and the Eye Disease Prevalence Research Group, 2004). One third of patients with vision threatening retinopathy will progress to legal blindness within 3 years if there is no intervention, however with timely intervention this figure could be reduced to one-tenth. These figures could be improved even further with new diagnostic technologies such as OCT and treatment options such as Vascular Endothelial Growth Factor (VEGF) blocking drugs and steroid treatments. These treatments are dependant on availability as there is a considerable cost to treating patients with an ongoing treatment plan. These data suggest the need for general population based screening for diabetes and diabetic retinopathy. The detection and treatment of diabetic eye disease in both the United States and Scandinavia is not
only cost effective, but is actually cost saving from a governmental perspective.

Approximately 350 million people have diabetes and diabetic retinopathy is one of the fastest growing reason for blindness, the other main reasons for blindness in western countries are age-related macular degeneration (AMD), glaucoma, retinopathy of prematurity (ROP) and in developing countries also cataract and trachoma. All these diseases can be screened and treated at their early stages (Danaei et al, 2011).

Comparably based on population sizes estimated savings in the United States could exceed 600 million USD annually, illustrating not only a reduction in vision loss but additional financial return on public funding. As mentioned earlier there are potentially many different effective ways of further improving on vision loss with drug interventions however these are not without cost. A comparison of early and deferred laser treatment with no treatment has proven that treatment for macular oedema with laser alone is extremely cost effective, based on 3-year outcomes (Ferris, 1991). Laser treatment has been shown to provide a gain of 0.236 in a quality-adjusted life year (QALY), a measure of disease burden that includes both the quality and the quantity of life lived (Mitchell et at, 2012). Such a gain was considered highly cost-effective relative to no treatment, and further cost saving can be found when it is used in combination with Vascular Endothelial Growth Factor (VEGF) blocking treatments such as ranibizumab (Mitchell et al 2012).

VEGF drugs themselves have also proven to be cost-effective. For example the RESTORE study data predicts a 0.26 QALY gain and an incremental cost-effectiveness ratio (ICER) of £10,412 for ranibizumab monotherapy relative to laser
therapy (Mitchell et al, 2012). In addition the cost-effectiveness is higher if the profile of patients being treated is younger, but this is dependent on the ability of a screening system being able to illustrate and refer the patient in a timely manner.

Software not only needs to be able to capture the retinal image of a patient but an assessor must be able to correctly identify and refer a condition. The detection of retinopathy is also somewhat dependent on the skills of the individual grader. The sensitivity and specificity of a grader is largely down to experience and systematic adherence to strict protocols. Automated grading of patients is not well established in developed countries. Even in programmes where automatic screening takes place the main use of such systems is to detect any disease verses no disease rather than to grade each case specifically.

An experienced grader is needed to make a final clinical judgement and so there will always be some degree of human grading used in screening programmes. Thus a grading support system is of more practical use than a fully automated screening service. The clarity of the retinal image is always of paramount importance as many diabetic patients suffer from cataracts, not only a natural manifestation of age, but also because of the early manifestation of cataracts in the diabetic population.

Images must not only clearly show the retinal vessels but also give an indication as to the various stages of retinopathy in order to channel patients to the correct referral pathways (see section 1.7). The more advanced the screening software the greater the chance of keeping the patient in an appropriate grading and treatment pathway. In the English diabetic eye screening programmes a large number of non-DR referable conditions are often detected, including wet age related macular
degeneration, optic disc cupping, vein and artery occlusions, detached retina, to name a few of the more urgent referable findings. In well-established DES Programmes, 50% of the urgent referrals have been for non-DR conditions.

Regular screening for both diabetes mellitus and diabetic eye disease should be the gold standard in preventing diabetic blindness globally. In the community of Laxa, County of Orebro, Sweden, such screening has been carried out since 1983. With the systematic screening of the population for diabetes and diabetic retinopathy, the loss of vision due to diabetic retinopathy is uncommon after 10 years (Olafsdottir and Stefansson, 2007). Zoega et al (2005) showed that there was also a significant relationship between screening compliance and visual outcome in diabetes patients in the Icelandic screening programme.

Loss of vision is catastrophic to any age group, particularly so in the working age group as it has personal, family and societal consequences. Studies have shown that early detection combined with appropriate treatment and management can prevent visual loss in up to 95% of these cases (Cheung and Wong, 2012). Diabetic retinopathy fulfils the WHO criteria for screening in that it evolves through key recognizable stages in the progression of blindness, represents an important public health problem, has a valid and acceptable screening tool of photographic screening and visual impairment can be slowed by effective and timely treatment (Andermann et al, 2008).

Visual impairment due to diabetic retinopathy does not become symptomatic until maculopathy or advanced proliferative disease occurs. So an effective annual screening programme would be expected to identify treatable retinopathy at a
preventable stage. The St Vincent declaration set out a target of reduction in new blindness due to DR by a third. The first service objective of the National Diabetic Retinopathy Screening Programmes in England and Wales was to achieve a reduction in blindness due to DR in the UK within 5 years by a minimum of 10 % with an achievable target of 40 % (Scanlon and Garvican, 2003). A near-comprehensive population coverage over 10 years have resulted in a decline in visual impairment due to diabetes (Gordon-Bennett, 2008) and DR is now as earlier mentioned no longer the commonest cause of blindness in the working age-group (Liew et al, 2014). Indeed, a recent survey in Cambridgeshire showed that DR is the fifth cause of visual impairment in persons of working age (16–64 years). Diabetic retinopathy is also a hallmark of other microvascular and macrovascular complications of diabetes. So screening for DR will enable timely monitoring and treatment of other diabetes induced morbidity.

1.5 Recent advances in diabetic retinal screening

The first diabetic retinal screening service in the UK was established in Cardiff. It was based on a basic ophthalmoscopic examination and was not improved upon until after a two-year study in Newcastle which trialled a non-mydriatic polaroid camera mounted in a mobile screening van. This van was driven to various locations to screen patients in the community (Taylor and Batey, 2007). One major finding of this study was that 1 in 20 patients could not be adequately photographed unless Tropicamide was used to dilate the patients pupils. This finding led to Tropicamide being used to achieve adequate dilation for retinal screening in England ever since.

The main finding of the study was that the photographic evaluation of Polaroid images was much better at detecting DR than that of ophthalmoscopy. This research
paved the way for the use of modern digital cameras as a means of screening for retinopathy. The findings from this study empowered the British Diabetic Association to attract charitable funding for ten more mobile screening units to cover other parts of the UK. In September 1994 the mobile screening vans had screened over 64,000 patients, 22% of whom received laser treatment as a result. Diabetic eye screening services (DES) have often been set up under a very tight budget with a very manual involvement from the staff. A simple grading pathway was often used and most software at the time was in its infancy.

Inevitably there are limitations to screening for diabetic retinopathy. Even the best programs will miss some cases of STDR. In 1984 at a Diabetes UK meeting standards were set out for a systematic screening program. These are often referred to as the ‘Exeter standards’, and have been widely accepted as standards for retinal screening and are featured in the NHS Diabetic Eye Screening Programme (DESP) guidelines (Curriculum.rcophth.ac.uk, 2014).

The Exeter standards cover different aspects of the screening methodology such as the sensitivity – the ability to detect pathology, and specificity - a measure of a test's effectiveness. The Exeter standards were based on a meta-analysis of the ability to effectively detect DR by different screening techniques, and various methods were gauged including ophthalmoscopy and digital photography throughout the 80s, 90s, and early 2000s against the ‘gold standard’ of a seven field stereo 35° photography and slit lamp biomicroscopy. This analysis found that overall only slit lamp biomicroscopy and digital photography achieved near 80% sensitivity and 95% specificity.
Digital fundus photography has become the preferred method for screening for DR. For best results mydriasis and two-field photography is generally recommended. Mydriasis reduces the technical failure rate from 27.1% to 8.3% using a single field and from 28.2% to 8.9% using two fields (Ding et al., 2012). Two-field strategy increases sensitivity from 75.6% to 87.8% without mydriasis and from 73.2% to 90.2% with mydriasis (Ding et al., 2012).

1.6 Photography

In England and Wales patients are required to have 4 images taken (in Scotland only a macular image of each eye is taken). The standard order of image capture is right macular, right nasal, left macular and then left nasal.

![Images](image_url)  
Right Macular view  Left Macular view  
Right Nasal view  Left Nasal view

Figure 8: This image shows the four images mandated by the English National Screening Programme for Diabetic Retinopathy (ENSPDR). In the majority of healthy patients these four images should be obtainable.

For each image the patient fixates on a target to minimise eye movements and to ensure that each image meets the national standard. Images can vary in quality due to
the final size of the pupil after dilation (in Scotland no dilation is used), the colour of
the retinal pigmentation, any forms of media opacity and/or any artefacts. A Retinal
 screener should if possible capture two nominal forty-five degree fields for each eye,
one centred on the fovea and one centred on the optic disc (Figure 8).

The quality of images taken at the time of screening is affected by a number of
variables, on the patient side age, mental health, physical health and willingness to
cooperate are some variables which can adversely affect screening outcomes. On the
side of the screener fatigue, stress, level of training and job satisfaction can play a
role. Training staff and in particular a screeners awareness of what constitutes an
acceptable/assessable image is vital to attaining a gradable set of images as eyelashes,
and mobility can directly affect the validity of the screening encounter.

![Figure 9](image)

**Figure 9:** Showing the definition of the macular. The Measurement Tool used by
Health Intelligence’s Spectra software is designed to allow graders to visually
compare distance based on the diameter of the optic disc. The tool will display
two circles centred at the fovea allowing the user to accurately see clinical
features within 1 disc diameter of the fovea and artefacts in the macular
Other considerations such as the software (see Figure 9), lighting of the rooms in which photography is performed, lighting in the room where grading is performed, colour gamut of monitors, types, sizes and resolution of displays, white balance settings on the camera, ISO speed settings, camera back types, lens barrel or pin-cushioning distortion, anti-aliasing filters on the camera chip, chip type such as CMOS vs. CCD and human judgment are just a few of the variables which will affect the ability of the grader to correctly identify referable features. The need for image clarity is paramount to achieving a gradable image on the NHS DESP guidance. If an image can be graded the outcome will determine the next stage for the patient, however if the image cannot be made clear then this will result in an onwards referral for a slit lamp bi-microscopy.

1.7 Grading Pathway & Outcomes

The English National Screening Programme for Diabetic Retinopathy (ENSPDR) set out standards for grading DR. Many screening programmes follow pathway 2 which was set out for non-Optometry based screening (see Figure 10). Pathway 2 involves the following approach:

- **Stage 1 First full disease grade (FFDG):** A grader accredited to do so, carries out a full disease grade on all image sets. Urgent referrals (R3) are immediately passed to the grading center for assessment by a Referral Outcome Grader (ROG).

- **Stage 2 Second full disease grade (SFDG):** A different grader will assess a random 10% of the no disease image sets and carry out a second full grade on all the disease image sets from the stage 1 grade. The Second Full Disease Grader does not see the result of the First Full Disease Grader prior to grading. Again urgent referrals
(R3) are passed to the grading centre for immediate assessment by a Referral Outcome Grader (ROG).

- Arbitration: If there is a difference of opinion about referral (sometimes referred to as a difference in outcome) between the two graders (FFDG and SFDG) then those image sets are referred onwards for an arbitration grade by a suitably accredited and experienced professional who will decide whether or not the patient should be referred to the hospital eye service (HES) or remain in the screening programme. Once again urgent referrals (R3) are passed to the grading centre for immediate assessment by a Referral Outcome Grader (ROG).

Figure 10: Showing the ‘Common pathway’. Detailing the flow of patients through screening to referral and re-invitation.
The purpose of the referral outcome grader (ROG) is to control the type of referrals (either DR or Non-DR) that are referred to the HES. Typically this will be carried out by the most experienced graders under the control of the programme clinical lead, or the ophthalmologist in charge of the HES that the patient is being referred to. The ideal set up is for the ROG grader to have full access to the patient’s clinical data however this is not always the case, unless the clinician has the time and access to the relevant data.

It is hoped that this type of screening pathway will allow screening programmes to better manage the inevitable ever increasing cost and workload they face. It is also worth noting that fatigue is a very real problem, with staff assuming that they are doing a good job while they can often be fatigued and miss pathology. Experienced grading staff should ideally grade no more than 80 cases a day in order to prevent fatigue and ensure that they are spending long enough on each case. It may well be that computer aided diagnosis can help in this regard.

1.8 Portable hand held retinal cameras

The recent development of lightweight portable cameras has driven innovation, allowing better retinal access to neurologists and paediatricians for monitoring of optic nerve swelling post neurosurgery, bed bound patients, and retinopathy of prematurity (ROP) screening for babies (see Figure 11). Probably the most exciting aspect of these cameras is the lower costs for the use in developing countries where telemedicine is essential to cover the vast geographical and socioeconomic challenges that hinder health care coverage.
Figure 11: Optomed Smart Scope Pro portable retinal camera. This clinician is taking an image of a child's eye with an optomed Smart Scope Pro portable retinal camera, this camera is one of the first which has been approved by the European Community (CE) and Federal Drugs Administration (FDA).

1.9 Online Databases for testing algorithms for automated screening purposes

There are many online databases associated with common eye conditions such as glaucoma, diabetic retinopathy and age-related macular degeneration. Generally the aim of such databases are to provide good examples of retinal images for research and training purposes. In particular this allows researchers to develop and compare different automated methods. Here we explore a few examples of currently available databases. These can all be found online and access to them is generally free for research purposes.

1.9.1 DRIVE database

Digital Retinal Images for Vessel Extraction (DRIVE) is a publicly available database (Staal et al., 2004; Isi.uu.nl,2014) consisting of forty colour fundus photographs has been used extensively to test many different automatic vessel extraction algorithms. The photographs were taken from a Diabetic Eye Screening Program (DESP) in the Netherlands. These images were chosen randomly from the retinal images of 453 subjects between 31 and 86 years of age. Each image has been JPEG compressed,
which is common practice in screening programs (see Figure 12). Seven of the forty images contain pathologies, such as exudates, haemorrhages and pigment epithelium changes. The forty images are divided into a test and training set each containing 20 images. Two qualified graders manually traced the retinal vasculature in these images to produce a mask of the retinal vasculature. In doing so the researchers manually segmented the retinal vessels from the underlying tissue. This process of segmentation, or extracting the pattern of blood vessels from the retinal tissue, is a task that numerous research groups have tried to automate as it is a laborious and time-consuming procedure.

Figure 12: Two example images from the DRIVE database. Image (a) shows a nasal view, (b) retina showing diabetic maculopathy. The images were taken with a Canon CR5 non-mydryatic 3-CCD camera with a 45 degree field of view (FOV). Each image is captured using 8 bits per colour plane at 768 × 584 pixels. The masked images have a resolution diameter of approximately 540 pixels.

1.9.2 STARE database

The STARE database (Hoover et al, 2000; Ces.clemson.edu, 2014) contains twenty digitized slides captured by a TopCon TRV-50 fundus camera taken at 35 degree field of view (see Figure 13). Ten of these contain DR pathology. All images were digitized to 605 × 700 pixels, 8 bits per colour channel. The approximate visible FOV is 650 × 500 pixels. Two image readers manually segmented all the images. The first
reader segmented 10.4% of pixels as vessels while the second segmented 14.9% of the pixels as vessels. The second observer segmented many more of the thinner vessels than the first one leading to this variation.

Figure 13: Two example 35 degree images from the STARE database: (a) healthy retina positions to show the optic disc in the nasal view, (b) pathological retina showing exudates in referable diabetic maculopathy.

1.9.3 ARIA online database

The ARIA online database was created in 2006 (Farnell et al, 2008; Aria-database.com, 2014), in a collaboration between St. Paul’s Eye Unit, Royal Liverpool University Hospital Trust, Liverpool, UK and the Department of Ophthalmology, Clinical Sciences, University of Liverpool, Liverpool, UK. The images were captured with a Zeiss FF450+ fundus camera at a 50 deg FOV and stored as uncompressed TIFF files.

Figure 14: These two control images taken from the ARIA online database, both have a resolution of 768 × 576 pixels in RGB with 8-bits per colour plane. The images show no referable pathology
This database is divided into three distinct groups: the first has 92 images with age-related macular degeneration, the second has 59 images with diabetes, and the third is a control group of 61 images (see Figure 14). The blood vessels, the optic disc and fovea locations were marked in each image by two image readers as a reference standard.

Figure 15: An example from the VICAVR database. Veins (marked in blue and white) and arteries (marked in red) are measured to compare vein to artery ratios.

1.9.4 VICAVR database

The VICAVR database consists of 58 images used for the computation of the A/V ratio (Ortega Hortas and Penas Centeno, 2010). The images have been captured on with a TopCon NW-100 model non mydriatic camera. All images are optic disc cantered with a resolution of 768 × 584 pixels (see Figure 15). The calibre of the vessels are measured at different radii from the optic disc and marked as artery or vein by three readers.

1.9.5 Messidor database

The Messidor database is the largest database with 1200 retinal images currently available on the internet (Agurto et al, 2010; Messidor.crhan.fr,2014). Three different ophthalmology departments were used to harvest images using a non-mydriatic 3CCD
camera fitted to a Topcon TRC NW6 at 45 degrees FOV. These were captured at a resolution of \(1440 \times 960, 2240 \times 1488\) or \(2304 \times 1536\) pixels. All images are stored in TIFF format (see Figure 16). 800 of these images were captured with pupil dilation. The reference standard provided with each image contains a diabetic retinal screening stage of progression of disease and the risk of macular oedema in each image.

Figure 16: Images taken from the Messidor database, the image on the left shows circinate rings of exudate with a registration mark (blue dot) highlighting the centre optic disc. The image on the right is the same image, however this time all exudates and haemorrhages have been marked to show progression.

Figure 17: The left image shows an image of diabetic maculopathy from the ImageRet database. The images on the right highlights areas of pathology marked by human graders, Image A showes hard exudates, B showes soft exudates, C shows soft exudates and D showes Heamorrhgaes. This type of image database is used) to test automated retinal grading software.
1.9.6 ImageRet database

Made publicly available in 2008 the ImageRet database is subdivided into two sub-databases, DIARETDB0 and DIARETDB1 (Valverde et al., 2010; it.lut.fi, 2014). The images were acquired with a reported 50 degree FOV using an unknown fundus camera with unknown settings at a size of $1500 \times 1152$ pixels in PNG format. DIARETDB0 contains 130 retinal images of which 20 were determined to be normal and 110 contain various signs of diabetic retinopathy (see Figure 17). DIARETDB1 contains 89 images, 5 images are of a healthy retina, 84 other images are reported to have some signs of mild proliferative diabetic retinopathy. The images were marked by four graders for the presence of microaneurysms, haemorrhages, and hard and soft exudates.

Figure 18: An image from the REVIEW database. This image shows a retinal image where the retinal blood vessels have been traced manually by human graders. This manual ‘segmentation’ reference allows for automated algorithms to be compared to the human segmentations.

1.9.7 REVIEW image set

The Retinal Vessel Image set for Estimation of Widths (REVIEW) database was made available online in 2008 by the University of Lincoln (Bashir et al, 2008). Sixteen mydriatic images with 193 annotated vessel segments consisting of 5066 profile points that were manually marked by three independent readers. Sixteen of the
images were subdivided into four sets, a high resolution image set (HRIS, 8 images), a vascular disease image set (VDIS, 4 images), a central light reflex image set (CLRIS, 2 images) and the kickpoint image set (KPIS, 2 images) (see Figure 18).

1.9.8 ROC database

Developed by the University of Iowa in 2009, the ROC microaneurysm dataset is a multi-year online database of microaneurysms (Niemeijer et al, 2010). 100 digital colour fundus photographs containing microaneurysms is divided in to two sets of 50 images test and non-test images. A reference standard indicating the location of the microaneurysms is provided with the training set (see Figure 19). The images were captured using a TopCon NW100 and a Canon CR5-45NM camera at 45 degree FOV and JPEG compressed. There are three different image sizes present in the database; 768 × 576, 1058 × 1061 and 1389 × 1383 pixels.

Figure 19 : An example retinal image from the ROC database. The righthand panel displays the boxed region of interest of the fundus image in the left panel. It is displayed in ‘red free’ mode with arrows pointing to microaneurysms.
1.9.9 The need for a realistic database

Considering the current databases available there still remains a need for a database that is more representative of the range of images collected in current DR screening programs. For example, the DRIVE database contains images that are comparable to screening services over a decade old. STARE uses scanned film images. For at least the last 10 years screening services in the UK have had to conform to certain types of camera, that have been approved by the national screening programme, together with suggested image resolutions and white balance settings. No two camera manufactures will use precisely the same optical pathway so variations will exist even within a single screening programme (Figure 20).

Even if images are restricted to the same camera and operator they still may show huge variations because of different settings, media opacities and even pupil diameters. We see this in screening services that have been in operation for many years. In Figure 20 we see two images of the same patient taken 2 years apart with different camera types. Although subtle the differences in camera type can change significant factors such as vessel calibre and bifurcation angles, this should be considered when automated systems are looking for specific features.

Figure 20: An example showing differences of white balance, exposure and barrel distortion. These images from the same patient images on 2 different cameras. A Topcon NW8 on the left and a Canon CGI on the right.
It is worth considering that a standard 100 ISO film image can hold up to the equivalent of 20 megapixels of information (Langford, 2000). The current range of cameras approved by the national screening programme still do not match the spatial resolution, dynamic range, or noise levels of the 35mm analogue 7 field stereo film photography such as that used in the UK Prospective Diabetes Study (1991).

All cameras currently used in retinal screening in England and Wales have been tested in accordance with the National Diabetic Eye Screening Programme guidance. Even though some discrepancies still exist between camera types there has been a movement towards standardisation from an operation point of view. With this in mind any image database specifically concerned with diabetic retinopathy should include cameras set up in line with the latest standards but also show the variation between camera types - ideally with a series of images from the same patient over time.

1.10 Current limitations with digital fundus images

The International Standards Organization (ISO) has a standardized scale for measuring the sensitivity of film to light. These standards have traditionally been used with film based fundus cameras. Screeners take images on a fundus camera where the exposure settings have been predetermined by software and the only variables the screener have are the possible small pupil settings or the flash level of the fundus camera. As camera technology has advanced, the ISO sensitivity settings have been changed, for example from 200 ISO to 400 ISO, so that the flash causes less discomfort for patients and in turn leads to a brighter image. The correct setting for use in Diabetic Eye Screening (DES) are published on the ENSPDR website. At a low ISO number a digital camera produces smoother and more detailed images than when
it is set to a high ISO, so the trade-off is that typically the flash will need to be high to achieve the right exposure for a higher resolution.

As mentioned the ISO standards followed are typically dictated by the image capture software of the fundus camera. These parameters are configurable for each individual patient, although typically they are seldom checked for consistency in eye screening programmes and may have been altered by the screener at the request of a patient for a less bright flash. Arguments for using a digital SLR camera typically outweigh those for film cameras, generally due to cost and convenience. However film is still very sensitive and can outperform that of current digital SLR cameras. Estimates of the resolution of a photograph taken with a 35mm film camera vary as more information may be recorded if fine-grain film is used in combination with a specialty film-developer. Conversely less resolution may be recorded with poor quality optics or with coarse-grained film (see Figure 21 for a comparison between digital and analogue film).

However the use of film in retinal screening would be problematic and scanning transparencies can also be problematic because of the film's tendency to scan with a blue colour cast. Some software producers deliver special colour profiles with their software to avoid this (Reuters, 2014).

The problem of dynamic range has been acknowledged by a number of digital single lens reflex (DSLR) manufacturers so for example some cameras have an automatic exposure bracketing mode that is used in conjunction with high dynamic range imaging software. Some sensors like the Fujifilm Super CCD combines sensors of different sizes to give increased dynamic range while other manufacturers use in-
camera software to prevent highlight overexposure such as the D-Lighting feature from Nikon.

![35 mm Film Diagram](image)

**Figure 21:** Showing the spatial resolution of film in units equivalent to digital camera megapixels. "Better" refers to spatial resolution only, the grey band is where the spatial resolution of film and digital are similar (adapted from Clarkvision.com, 2014).

It is worth noting that the sensitivity of the retinal screening programme is solely dependent on the ability of a screener to be able to see pathology in the image of the retina. Unlike pilots and police drivers there is no official guidance specifying regular vision check-ups for a retinal screener, no minimum standard, and no checks take place on the graders ability to visualise images or tests for colour vision defects. In today’s diabetic eye screening environment it would be ideal for a computer to grade automatically 24 hours a day 7 days a week with high sensitivity and specificity. However the current recommendation is that a trained and accredited grader should be used until the automatic detection software has developed further and it is capable of identifying securely a good range of clinically significant data.
Figure 22 below is a stark reminder to how much a missed sight threatening pathology can affect a patient and also the effect to a screening programme if a patient was to successfully pursue a claim of clinical malpractice. For a private company to even consider fully automated retinal screening as an option in place of a team of graders the evidence would need to be compelling.

<table>
<thead>
<tr>
<th>Injuries Affecting Senses</th>
<th>Estimated Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Blindness</td>
<td>£155,250 to £500,000</td>
</tr>
<tr>
<td>Loss of Sight in One Eye with Reduced Vision in the Remaining Eye</td>
<td>£37,150 to £104,500</td>
</tr>
<tr>
<td>Total Loss of One Eye</td>
<td>£32,000 to £38,175</td>
</tr>
<tr>
<td>Complete Loss of Sight in One Eye</td>
<td>£28,750 to £32,000</td>
</tr>
<tr>
<td>Cases of serious but incomplete loss of vision in one eye without significant risk of loss or reduction of vision in the remaining eye, or where there is constant double vision</td>
<td>£13,750 to £22,650</td>
</tr>
<tr>
<td>Minor but permanent impairment of vision in one eye, including cases where there is some double vision, which may not be constant</td>
<td>£7,375 to £12,200</td>
</tr>
</tbody>
</table>

Figure 22: Estimated costs of clinical malpractice resulting from misdiagnosis (trinitylaw.co.uk, 2015)
In 2002 twin boys won £1m in damages after a hospital failed to spot retinopathy of prematurity at birth (News.bbc.co.uk, 2002), and clear viewing of the retina is essential in determining sight threatening pathology. Subtle discrepancies in image clarity can greatly affect the outcome when trying to visualise the retina. For example the lighting at the retinal screening location may affect the ability of the screener to detect pathology (florescent versus incandescent ambient lighting). These varying light conditions mean that the screener’s eyes may be affected and light adaptation may effect colour and luminosity perception.

Figure 23: Effects of image manipulation. A) Displays an original image, alongside examples of the kinds of inappropriate manipulation that are common with inexperienced screeners. B) displays the same image with typical over enhancement. C) displays a red free filter image which increases the contrast of the original image but distorts the image. D) displays the original image of a self-sclerosed microaneurysm. E) displays a typical over manipulation of the image (D) that makes feature look like an exudate, and F) shows the correct amount and type of manipulation that exhibits a subtle grey representation true to the real likeness of the pathology.

The clarity and assessibility of an image is rather subjective and national standards statements such as ‘fine vessels on the disc’ can be misinterpreted by different
screeners. Even where groups of specialists form expert panels there is often high levels of variability in interpretation. One common reason for poor images is merely that the flash settings were set too low for the photography to show the retina in any clarity. This may highlight issues within the fundus camera such as a weakening capacitor or incorrect base level flash setting which needs changing in the engineer’s flash setting menu.

Current methods used in retinal screening to enhance images involve the use of brightness and contrast. However the over use of these controls can often make features disappear by posterization ‘blowing out’ detail. This can be seen in Figure 23 where distortion can transform a normal vessel (Figure 23(A)) to appear as pathology such as intra retinal microvascular abnormalities (IRMA) and therefore lead to false positive results (Figure 23(B)). This phenomenon has also been noticed by DESP. New guidance that came into effect in April 2013 which states that unless IRMA is spotted in the colour image it should not be graded as being present.

![Image showing colour distortions. (A)The image on the left shows an original unenhanced image of the optic disc, the image on the right (B) shows the distortions that occur at the disc when brightness and contrast is applied to the original image.](image)
Figure 25: Retinal images showing a microaneurysm. A microaneurysm (marked by an oval) is present in A and C (the original image). B is result of imaging the same location but under exposing the retina and D has started to become posterised as the brightness and contrast has been applied solely to enhance the image to the point where the microaneurysm is visible.

The optic disc often displays a large difference in luminance from the background tissue and is commonly affected by manipulation. In addition this area can harbour new vessels in diabetic patients, peripapillary membranes in high myopic patients, glaucomatous changes and indications of potentially life threatening raised intracranial pressure. So correct exposure and visualisation is essential when viewing retinal images, as all of these optic disc features can be masked when achieving correct brightness levels post manipulation.
1.11 Gauging assessablity of retinal images

As mentioned previously in English screening programs, a screener should capture two nominally forty-five degree fields for each eye, one centred on the fovea and one centred on the optic disc. Extra images may be required, for example when there is noticeable media opacity such as a cataract, as an anterior image may be useful in illustrating the reason for a blurred image and rule out suspicions of the image merely being taken out of focus or even worse that there may be a vitreous haemorrhage so dense that little or no light is passing through. There may also be a feature on the anterior of the eye such a Pterygium. The National screening committee (NSC) guidance state that failure to meet definition of adequate results in an ungradable image. This is unless referable diabetic retinopathy (R2, R3, M1, unstable treated proliferative diabetic retinopathy) is visible anywhere in the images. The NSC guidance also states that in some unusual cases such as patients with a large disc an image may fall within both good and adequate categories and in such cases the image should be classified as good.

For some diabetic patients, digital retinal photography does not generate retinal images deemed to be ‘adequate’ by retinal graders. Images that are unassessable for retinal grading can be unassessable for a number of reasons.

A significant proportion of ungradable images are a result of cataracts or opacity caused as a result of cataract surgery (see Figure 26). Cataracts are more common in diabetic patients and the process can somewhat be accelerated because of the way that the sorbital-aldose reductase pathway effects the metabolism of sorbital. In addition the fundus can vary in colour due to differences in the retinal pigmented epithelium between individuals. For example the retina may appear green
or brown in Asian and Afro-Caribbean eyes. This difference in coloration coupled with a media opacity can hinder a grader's ability to detect retinal pathology as the images may appear blurred or murky.

Figure 26: Showing the progression of cataract in a diabetic patient taken over a 5 year period (2006-10). This clearly demonstrates the progressive deterioration in the quality of images as a cataract develops. In 2010 the patient had cataract surgery and the quality of the retinal image increases significantly.

Retinal graders are continuously monitored for quality and reliability via a monthly online quality assurance exam. Graders are recommended to take these at least 6 times
a year to demonstrate competency. However there is no option to grade an image as being unassessable on these monthly tests. So staff vary greatly in their perception of what is assessable. The burden on graders to examine degraded images would be greatly reduced by the introduction of a software tool that automatically rated the assessability of an image and indicated when images may be deemed unassessable (or inadequate).

Figure 27: Example images from the Scottish grading scheme. A) displays a red free image with a clear view of the retina. The center of the fovea is greater than two disk diameters from the edge and vessels are clearly visible within one disk diameter of the fovea. B) displays a retinal image with moderate media opacity while C) shows a fundus image from a patient with a dense cataract. D) shows all the vessels visible within 1 DD of the centre of the fovea from (A). E) shows a moderate view of the vessels within 1DD of (B) and finally F) shows virtually no details of retinal vessels can be made out within 1DD of the fovea of the patient with the cataract.

Sometimes it can be difficult for a grader to be able to accurately gauge the difference between a gradable and ungradable image as it typically depends on the graders skill at image manipulation (see Figure 27 for examples from the Scottish
guidance). A high unassessable rate within a screening programme would put a higher load on slit lamp bio-microscopy (SLBM) clinics leading to a greater delay for other patients and extra cost to the service. Patients who are reviewed at SLBM may also be re-invited for SLBM, and so these unassessable patients may technically change to partially assessable if year on year they are now invited for SLBM. Unless the optometrist also takes an image of his patient to gauge the assessibility from photographic means he may not ever know if this patient was invited for SLBM just because of a poor quality image.

Figure 28: Screening outcomes for patients in East Anglia Eye Screening Programme (2012-2013). 75,663 patients were screened and 3% were sent for further examination by slit lamp, these patients were sent there as there was no clear view of the retina in the images taken at the time of screening.

Another parameter that is important to grading accuracy is time. The fact is that some graders will spend a lot more time than others trying to make a lower quality retinal image accessible by patiently manipulating contrast and luminance. It would be a huge benefit to screening programmes in general if graders were as consistent as possible with their application and manipulation of software filters. For example if a
filter was available to staff to standardize manipulation at the press of a button no
doubt there would also be substantial time saving enabling staff to grade more images
while becoming less fatigued. A more regulated method of image analysis may put
less strain on the HES. At multi-disciplinary team meetings nationally over-
manipulation of retinal images is often cited as a cause for concern and staff are often
encouraged not to manipulate image parameters.

1.12 Possible solution: The cataract filter algorithm (Clearvue)

The amount of scattering in a young non pathological eye is low but will increase
with age (e.g. cataract) (Hennelly et al, 2002). This increased scattering will result in a
degradation of the retinal image in terms of reduced brightness, poorer contrast,
colour degradation, and possibly lower spatial resolution. Such a problem is similar to
degradation of images by the atmosphere (see Figure 29).

**Figure 29:** Images depicting the similarities between pictures taken on a foggy
day and fundus images taken through a media opacity. There are similar
difficulties in making out the features of the branches of the trees and the
branches of retinal arteries and veins. Both images exhibit the same inability to
see fine details which is caused not by poor focus but by light scatter.

Although a large body of work exist on the enhancement of images recorded in poor
visibility conditions, a paucity of papers have focused on reducing the degradation of
retinal images by intraocular scattering (Peli and Peli, 1989). Image contrast
enhancement can be divided into two main methods, model based and non-model
based. Model based algorithms improve image contrast by reversing the underlying cause of image degradation but requires a large amount of information about the geometry of the scene and the nature of the scattering. It is difficulty to develop a realistic yet tractable model of intraocular scattering as many factors can contribute to the scattering of light in a fundus image. Perhaps the most important non-model based algorithm is histogram equalization. This algorithm is usually performed on the luminance and saturation of the image but not the hue so as to maintain the original colour. However, because intraocular scattering is partly wavelength dependent (Costello et al, 2007), this method may cause some colour degradation.

Oakley and Bu in 2007 presented a successful non-model based algorithm to restore the chrominance and luminance of a scene while maintaining good colour fidelity. In this model, the image (I) recorded by the camera is the sum of the light reflected by the scene (here the retina (R)) as it would be observed in the absence of scattering together with the scattered light (S), known in the remote sensing literature as “airlight”. The airlight value will depend on the distribution and the concentration of the different scattering particles (e.g. keratocytes in the cornea, proteins aggregates in the crystalline lens), the distance from the eye to the camera and the angle of illumination. As a result, the airlight value can vary across the image.

Figure 30: An airborne image degraded by haze. The image on the left is an example image captured by an airbourne camera which has been degraded by
haze. The image on the right demonstrates how the visual haze can be effectively filtered out to increase visibility (Oakley and Bu, 2007).

This system could be potentially adapted for the purposes of retinal screening. Figure 30 displays an airborne image degraded by atmospheric haze which demonstrates a large variation in the range from the foreground to the background. This results in a non-uniform loss of contrast across the image that in many ways mimics the effect of a media opacity in a retinal image. Figure 30 also displays the output from the Clearvue algorithm which manages to remove most of the haze from the original image.

1.13 Overview of the rest of the thesis

As detailed in section 1.9.9 there exists the need for a database of fundus images that is more representative of the images taken from diabetic patients in current screening programs. Chapter 2 outlines the efforts to develop such a database. As many diabetic patients have comorbidities we have grouped our images based on these conditions and have applied two automatic vessel extraction algorithms to the database (see Appendix A for a preprint of our manuscript (Holm et al., 2014). As part of this database development, vessel masks that distinguish the vascular bed from the surrounding tissue needed to be expertly drawn by hand. As described in chapter 2, three different semi-automatic methods were trialled by the author and a medical retina ophthalmologist and the best one employed to help generate the vessel masks. Each mask, which is associated with a single fundus image, forms part of the database as the ideal solution to the problem of vessel segmentation.

Following on from the development of the diabetic retinopathy database, chapter 3 details our initial steps into improving the “assessability” of fundus images that have previously been deemed inadequate. These preliminary experiments
involved various tests of the Clearvue ‘cataract filter’. In particular tests were carried out that initially involved investigating the effects of the two variables associated with the cataract filter on the clarity and quality of the processed images. The "dark level" parameter is used to mitigate the effect/extent of noise in the input image and ranges from 0 to 2. The level of “smoothness” is set by the second control parameter (the scale) which ranges from 1-15%. This limits the scale over which the enhancement is allowed to adapt according to the local image properties. The higher the scale the ‘smoother’ the coverage and the less local the effect of filtering (the broader the spatial filter).

Once an agreement was reached between the author and a medical retinal ophthalmologist regarding the best parameters to use a sequence of 100 previously inadequate images were reprocessed through the cataract filter and re-evaluated by the author and the medical retinal ophthalmologist. This lead to 12 cases being deemed adequate for grading with some referable cases being noted. An earlier pilot part of this study has been published (Russell et al., 2012) and is included in Appendix B.

Finally in chapter 4 the limitations of our initial study are listed and suggestions are made for future work that may improve performance and areas for future investigation.
Chapter 2: DR HAGIS – A new realistic DR Database for the Automatic Extraction of Retinal Surface Vessels

2.1 The need for a realistic DR database

As detailed in Chapter 1 there already exists a number of publically available databases containing retinal fundus images. Some of these are specifically designed to test automatic vessel extraction algorithms (Hoover et al., 2000; Staal et al., 2004; Farnell et al., 2008). However the images in these databases do not accurately represent the quality, complexity, and resolution that any generic automatic image analysis system will have to address in a modern day screening programme. Specifically in all the publically available databases the images are of much lower resolution than modern digital cameras, are taken in idealized imaging conditions, with the same camera and operator. While these databases have been particularly useful in allowing researcher to compare for example automatic vessel extraction algorithms they do not represent what a modern day grader will typically see. In addition patients with diabetes often have co-morbidities. To address the lack of a realistic, publically available, database of fundus images we decided to generate the DR HAGIS (Diabetic Retinopathy, Hypertension, Age-related macular degeneration and Glaucoma ImageS) database. The main aim of this chapter is to report on the development of a new publically available realistic database of DR fundus images (Dr HAGIS). These fundus images were segmented by hand by the author and used by collaborators to test two state-of-the-art automatic segmentation algorithms.

2.2 DR HAGIS database

This database consists of thirty-nine high-resolution images provided by Health
Intelligence (Sandbach, UK). These were taken from diabetic patients attending a DR screening programme run by Health Intelligence and were recorded from different DR screening centers in the UK. All patients gave ethical approval for the use of these images for medical research. The thirty-nine images were grouped into one of four co-morbidity subgroups: glaucoma (images 1-10), hypertension (images 11-20), DR (images 21-30) and AMD (images 31-40). One image was grouped into two subgroups, as this patient was diagnosed with both AMD and DR (images 24 and 32 are identical). A total of three different non-mydriatic fundus cameras were used to capture the fundus images: Canon CR DGi (Canon Inc, Tokyo, Japan), Topcon TRC-NW6s (Topcon Medical Systems, Oakland, NJ), and Topcon TRC-NW8 (Topcon Medical Systems, Oakland, NJ). All fundus images have a horizontal FOV of 45 deg. Depending on the digital camera used, the images have a resolution of 4752x3168 pixels, 3456x2304 pixels, 3216x2136 pixels, 2896x1944 pixels, or 2816x1880 pixels.

2.3 Ground truth images

Ground truth images corresponding to the segmented vessel patterns were generated using GIMP Image Manipulation software (Gimp.org, 2014). The original images were first opened in GIMP as Jpeg files. Then transparent layers were overlaid on each original image to try a number of methods to determine which gives the most accurate two-bit black and white tracing of the vessels.

The first method tested was to make the image monochromatic, threshold the histogram level until maximum contrast level had been reached to assist in the segmentation, any unwanted pixels were erased and then any gaps between the vessels walls were filled-in. Unfortunately, this method proved to be both time consuming and inconstant, with too much variably when comparing the transparency
layer to original image (Figure 31 displays one example of this approach).

Figure 31: An example of segmentation method one. This figure displays both the traced image attained by thresholding a binary version of the original image and then filling in between blood vessels (left) versus the original image (right). The traced image was less clear than the original image and as such this was not a suitable segmentation method.

A second method tested was to trace the path of each vessel in sections. Each section was processed separately and then pieced together to form the final mask. The final map was manually corrected with the eraser tool, however once again this proved to be time consuming taking on average three hours to complete a single retinal image (see Figure 32 for an example of this method).

The final (and most effective) method used was to utilize the line tool to trace each vessel and to manually increase and decrease the brush size until the vessel diameter was matched. The brush tool in GIMP displays a target ring that is visible above the vessel and so gives a visual guide to the caliber of the brush tip. This method facilitated drawing a much smoother and solid continual line, this technique proved to be both visually accurate and rapid taking on average 40 minutes per image. Throughout the process of tracing the drawn layer was turned on and off manually.
and rapidly to allow for the accurate checking and rechecking of line width, this technique could be likened to manually flicking between cells in animation to gauge if any change had taken place. Finally any imperfections were then erased pixel by pixel with the eraser tool, the end result was a relatively accurate and smooth representation of the underlying retinal vessels. This technique was then applied to all images.

Figure 32: Showing a manually traced image in sections. In this image seen as dots (left side of image) which could be manipulated later to change the angle and then stroked (painted) to desired vessel calibre (right), this method however proved to be inadequate for segmentation.

Figure 33: Showing an example of the final image tracing technique. As seen on the image on the right, this 2 bit tracing appears smooth and well defined and was by far the most effective method to segment retinal images.
2.4 Initial retinal vessel segmentation results

We applied two previously published automatic vessel extraction algorithms (Holm and McLoughlin, 2014) to this new database to generate some initial segmentation results (see Appendix A for pre-print of our submitted paper). One approach was based purely on the intensity of the pixels in the fundus image (intensity-based) while the other is based on the output of oriented Gabor filters (Daugman, 1985) which preprocessed each fundus image.

The Gabor filter approach roughly follows that of Rangayyan et al. (2007). In short, for each fundus image in the database a background estimate was subtracted from the green-channel image. Then twelve differently oriented Gabor filters, differing in orientation by 15 degrees, were applied to the field of view (FOV). All pixels outside the FOV were set to zero. Next a segmentation threshold was applied to the filter responses to generate binary masks of the vasculature. When more than one scale of Gabor filters was employed the binary masks of the vasculature were combined together to form a single segmented image. Finally post-processing steps were used to reduce as much as possible any edge effects of the FOV (see Figure 34 for an example of this approach).

The intensity-based approach was largely based on the vessel extraction algorithm of Saleh et al. (2011). Basically this approach includes several steps aimed at reducing the noise and illumination variation across the fundus images, thereby increasing the contrast of the vasculature. Top hat and bottom hat filters were used to reduce the effects of uneven contrast within the fundus images. Next the background illumination was removed by subtracting the output of a large median filtered version of each image from itself. This reduces variations in the background intensities. Then
after Gaussian smoothing, a h-maxima transform was applied to segment the vascular bed with the threshold defined as in Saleh et al. (2011). Once again post-processing steps were used to reduce any edge effects of the FOV (see Figure 35 for an example of this approach).

Figure 34: Example fundus images from the Dr Hagis database with automatic segmentation using the intensity based approach. (A) This is an image of a patient suffering from DR. (B) The results of the vessel extraction using the intensity based approach. Green pixels highlight correctly segmented pixels (true positives), red pixels show those vessel pixels false segmented as background (false negatives), blue pixels are the oversegmented (false positives), and black pixels represent correctly segmented background pixels (true negatives).

The quality of segmentation was determined by the mean percentage of correctly segmented pixels within the FOV. The FOV is defined by the provided mask images. Sensitivity is defined as the percentage of vessel pixels within the FOV segmented as much, and the specificity as the percentage of correctly segmented
background pixels (again within the FOV). Finally the accuracy is the percentage of correctly segmented pixels (whether vessel or background).

Figure 35: Example fundus images from the Dr Hagis database with automatic segmentation using a two scale gabor filter based approach. (A) The same fundus image from Figure 34 showing an image of a patient suffering from DR. (B) The results of the vessel extraction using the two scale Gabor filter approach. Color coding is identical to Figure 34.

Table 1 presents the results from both the intensity based and two scale Gabor filter approaches. Results are presented as mean (± standard deviation). Both the approaches resulted in similar overall segmentation accuracy (intensity based: 95.83%, 2-scale Gabor filter approach: 95.71%). For the DRIVE database, the degree of correspondence between two expert observers was 94.73% (Staal et al., 2004). So if a similar inter-observer accuracy is assumed we can conclude that both approaches perform as well as an expert human observer in terms of overall segmentation accuracy.
Table 1: Accuracy, Sensitivity and Specificity for the intensity based approach and Gabor approach.

<table>
<thead>
<tr>
<th>Intensity Based Approach</th>
<th></th>
<th></th>
<th></th>
<th>Two Scale Gabor Filter Approach</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>Accuracy</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Image</td>
<td>Accuracy</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>1</td>
<td>97.03%</td>
<td>66.73%</td>
<td>98.95%</td>
<td>21</td>
<td>95.43%</td>
<td>59.77%</td>
<td>98.95%</td>
</tr>
<tr>
<td>2</td>
<td>95.70%</td>
<td>62.32%</td>
<td>98.46%</td>
<td>22</td>
<td>95.69%</td>
<td>61.30%</td>
<td>98.68%</td>
</tr>
<tr>
<td>3</td>
<td>95.97%</td>
<td>53.37%</td>
<td>99.34%</td>
<td>23</td>
<td>94.29%</td>
<td>50.10%</td>
<td>98.83%</td>
</tr>
<tr>
<td>4</td>
<td>95.54%</td>
<td>53.74%</td>
<td>99.15%</td>
<td>24</td>
<td>95.22%</td>
<td>48.61%</td>
<td>98.36%</td>
</tr>
<tr>
<td>5</td>
<td>95.89%</td>
<td>57.50%</td>
<td>98.70%</td>
<td>25</td>
<td>95.90%</td>
<td>52.23%</td>
<td>98.48%</td>
</tr>
<tr>
<td>6</td>
<td>96.38%</td>
<td>49.80%</td>
<td>98.68%</td>
<td>26</td>
<td>95.43%</td>
<td>50.12%</td>
<td>98.73%</td>
</tr>
<tr>
<td>7</td>
<td>95.53%</td>
<td>46.73%</td>
<td>99.34%</td>
<td>27</td>
<td>94.71%</td>
<td>52.68%</td>
<td>98.31%</td>
</tr>
<tr>
<td>8</td>
<td>96.03%</td>
<td>59.40%</td>
<td>98.86%</td>
<td>28</td>
<td>96.45%</td>
<td>58.02%</td>
<td>98.53%</td>
</tr>
<tr>
<td>9</td>
<td>96.47%</td>
<td>65.38%</td>
<td>98.89%</td>
<td>29</td>
<td>95.75%</td>
<td>61.06%</td>
<td>98.81%</td>
</tr>
<tr>
<td>10</td>
<td>96.18%</td>
<td>59.43%</td>
<td>99.37%</td>
<td>30</td>
<td>95.37%</td>
<td>49.90%</td>
<td>99.51%</td>
</tr>
<tr>
<td>11</td>
<td>96.36%</td>
<td>71.45%</td>
<td>98.90%</td>
<td>31</td>
<td>95.06%</td>
<td>59.89%</td>
<td>99.04%</td>
</tr>
<tr>
<td>12</td>
<td>97.06%</td>
<td>62.18%</td>
<td>98.65%</td>
<td>32</td>
<td>95.58%</td>
<td>49.42%</td>
<td>99.07%</td>
</tr>
<tr>
<td>13</td>
<td>97.26%</td>
<td>56.42%</td>
<td>98.28%</td>
<td>33</td>
<td>95.24%</td>
<td>55.93%</td>
<td>98.51%</td>
</tr>
<tr>
<td>14</td>
<td>96.22%</td>
<td>63.41%</td>
<td>99.09%</td>
<td>34</td>
<td>96.50%</td>
<td>49.70%</td>
<td>98.95%</td>
</tr>
<tr>
<td>15</td>
<td>96.15%</td>
<td>50.43%</td>
<td>99.61%</td>
<td>35</td>
<td>95.32%</td>
<td>52.86%</td>
<td>99.20%</td>
</tr>
<tr>
<td>16</td>
<td>96.44%</td>
<td>51.75%</td>
<td>99.38%</td>
<td>36</td>
<td>96.25%</td>
<td>58.07%</td>
<td>98.92%</td>
</tr>
<tr>
<td>17</td>
<td>95.55%</td>
<td>67.44%</td>
<td>98.36%</td>
<td>37</td>
<td>95.25%</td>
<td>43.57%</td>
<td>99.25%</td>
</tr>
<tr>
<td>18</td>
<td>95.94%</td>
<td>56.75%</td>
<td>99.14%</td>
<td>38</td>
<td>95.97%</td>
<td>58.94%</td>
<td>98.65%</td>
</tr>
<tr>
<td>19</td>
<td>96.70%</td>
<td>50.93%</td>
<td>98.82%</td>
<td>39</td>
<td>95.02%</td>
<td>53.14%</td>
<td>99.09%</td>
</tr>
<tr>
<td>20</td>
<td>94.65%</td>
<td>47.07%</td>
<td>99.47%</td>
<td>MEAN</td>
<td>95.83%</td>
<td>58.83%</td>
<td>98.91%</td>
</tr>
<tr>
<td>21</td>
<td>95.43%</td>
<td>59.77%</td>
<td>98.95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>95.69%</td>
<td>61.30%</td>
<td>98.68%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>94.29%</td>
<td>50.10%</td>
<td>98.83%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>95.22%</td>
<td>48.61%</td>
<td>98.36%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>95.90%</td>
<td>52.23%</td>
<td>98.48%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>95.43%</td>
<td>50.12%</td>
<td>98.73%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>94.71%</td>
<td>52.68%</td>
<td>98.31%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>96.45%</td>
<td>58.02%</td>
<td>98.53%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>95.75%</td>
<td>61.06%</td>
<td>98.81%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>95.37%</td>
<td>49.90%</td>
<td>99.51%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>95.06%</td>
<td>59.89%</td>
<td>99.04%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>95.58%</td>
<td>49.42%</td>
<td>99.07%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>95.24%</td>
<td>55.93%</td>
<td>98.51%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>96.50%</td>
<td>49.70%</td>
<td>98.95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>95.32%</td>
<td>52.86%</td>
<td>99.20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>96.25%</td>
<td>58.07%</td>
<td>98.92%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>95.25%</td>
<td>43.57%</td>
<td>99.25%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>95.97%</td>
<td>58.94%</td>
<td>98.65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>95.02%</td>
<td>53.14%</td>
<td>99.09%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>94.65%</td>
<td>47.07%</td>
<td>99.47%</td>
<td>MEAN</td>
<td>95.83%</td>
<td>58.83%</td>
<td>98.91%</td>
</tr>
</tbody>
</table>

Mean values for Intensity Based Approach: Accuracy 95.83%, Sensitivity 58.83%, Specificity 98.91%.
2.5 Discussion

Developments particularly in the resolution of fundus cameras in the last decade or so have vastly changed the quality and size of digital fundus images now routinely captured in NHS DR screening centers. Currently there is no comparable high-resolution database of realistic DR images available to test automatic vessel extraction/segmentation algorithms on. The DR HAGIS database fills this gap and provides a realistic collection of high-resolution fundus images collected in the current NHS screening environment.

The vascular maps, generated by the author, are as close as possible (with current technologies) to defining what an expert human sees when observing digital images of the retina. Each vascular map was painstakingly created to ensure as accurate a segmentation as possible. Each vascular map took over an hour to carefully trace.

Table 1 presents our initial results for the intensity based and two-scale Gabor approaches compared against the manually segmented maps. Both approaches resulted in similar overall segmentation accuracy (intensity based: 95.83%, 2-scale Gabor filter approach: 95.71%). These algorithms, which represent two current approaches to retinal vessel segmentation, do a reasonable job in successfully extracting details of the larger retinal vessels but they clearly leave room for improvement as their sensitivity is only 55.83% and 59.71%. That said the main aim of this work was to generate a database of realistic DR fundus images that other research groups can test their innovations on.
Chapter 3: Filtering the effects of cataracts

3.1 Background

Previous guidance on assessability of retinal fundus images specified a perceived ‘cut off’ defining a minimum limit of clarity. The automatic detection of the third generation retinal vessels should make it possible to identify whether an image is assessable and this may be beneficial in reducing the numbers of false negative results. However an initial test carried out by Health Intelligence showed a general lack of agreement between experts on what is adequate/assessable. As part of an audit on assessability we had a group of 18 graders examine 26 images all of which were originally graded as unassessable (one example is shown in figure 35). Despite all these images being previously graded as unassessable 11 of 18 (61%) graders felt that over half of the images were assessable.

Figure 36: An example of an unassessable retinal image. This image was graded as assessable by a retinal grader, however the fine vessels within 1DD of the centre of the fovea are difficult to visualise within this image. This hazy view inhibits a clear view of the retina and thus the image is now classified as ‘inadequate’ by the ENSPDR guidance.
This illustrates the subjective nature of retinal assessment, what is assessable to a clinician may not necessarily be assessable to a grader and vice versa. Complaints have been voiced regarding the assessability of some of the grading image sets available on the online national EQA grading system. Graders have complained that they have been forced to grade unassessable images as part of their monthly test despite the fact that all these images were deemed assessable.

3.2 Current image enhancement techniques to improve image quality

There are four main factors which effect image quality colour, focus, contrast and illumination. Some images will need adjustment of brightness and contrast because of media opacities, some will require adjustment as the image may be under exposed. Current methods used in retinal screening to enhance images involve the use of brightness and contrast, however as described in the introduction the over use of these controls can often make features disappear by posterization or ‘blowing out’ of detail (figure 23). Distortion can for example make fine vessels appear as pathology such as intra retinal microvascular abnormalities (IRMA) and therefore lead to false positive results. This phenomenon has also been noticed by DESP and the latest guidance states that unless IRMA is spotted in the colour image it should not be graded as present.

One simple modification which shows promise in increasing sensitivity of fundus cameras is the removal of the Anti-Aliasing filter (AA). The AA filter is commonly placed on the surface of the image sensor on a digital single lens reflex (SLR) camera. These filters are designed to deal with fine repeating patterns, as fine or finer than the pitch of the actual pixels. Such patterns are not common place in the retina and an anti-aliasing filter works by blurring the image just a tiny amount (only
fractions of pixels) so that no details finer than the distance from one pixel to the next hit the sensor.

Removal of this anti-aliasing filter gives a much sharper image at high levels of zoom, when examining small clusters of pixels. However this is not the case on any of the currently approved SLR cameras used in DRS screening although this may change in the near future when cameras like the Nikon 800e come on to the market.

3.3 Methods

As described previously the amount of scattering in a young non pathological eye is low but will increase with age and some pathology (e.g. cataract). This increased scattering will result in a degradation of the retinal image similar to degradation of images by the atmosphere. Although a large body of work exists on the enhancement of images recorded in poor visibility conditions, less research has been carried out on how to reduce the degradation of retinal images by intraocular scattering.

One possible approach is based on research carried out in the Faculty of Engineering and Physical Sciences at the University of Manchester who developed a powerful filter to mitigate contrast loss due to atmospheric scattering effects. This algorithm was made commercially available in both hardware and software formats by a spinout company Dmist Research Ltd. under the brand names “Clearvue” and “ClearvueHD”. In principle the same technology could be applied to mitigate the scattering effects of media opacities such as cataracts. To test the potential of this Clearvue filter in retinal screening a commercial agreement was put in place between Health Intelligence Ltd and Dmist Research Ltd. We refer to this as the ‘cataract filter’ in the context of retinal screening.
The cataract filter is based on simple subtraction. The image is first converted to a linear intensity scale. Images are conventionally represented using gamma encoding. RGB pixel values do not represent intensity but (roughly) the square root of intensity. An estimate for the scattered component is then generated and is then subtracted from the input image. The scattered component usually varies in different parts of the image. The cataract filter represents this component as an arbitrary smooth function. The level of “smoothness” is variable according to an input parameter (filter scale). The output is then rescaled and the non-linear (gamma) encoding is re-applied.

\[ CV = \frac{p_k - \bar{p}_k}{\bar{p}_k} \]

Figure 37: Example of an artificial scene in simulated clear and foggy conditions. Where \( p_k \) represents the pixel intensity and \( \bar{p}_k \) represents the same pixel in a low-pass filtered version of the image. \( \frac{p_k - \bar{p}_k}{\bar{p}_k} \) gives a measure of local contrast within the image namely the coefficient of variation (CV).

The cataract filter is designed to mitigate contrast loss due to scattering effects, the image model works based on the principles outlined in figure 37. Foggy scenes exhibit a lower coefficient of variation and we can use this statistical signature of the fogginess to determine an estimate of the additive scattering component. The cataract
filter achieves this using a iterative process designed to make the CV in the image as uniform as possible.

This method is designed to overcome these faults of colour and feature distortion, allowing the colour levels and contrast to be optimised automatically to potentially offer retinal graders additional information. If the cataract filter works as is hoped it should help reduce the rate of ungradable/unassessable images (the current percentage of unassessable images for most screening programmes is higher than the 7% recommended rate by the DESP) and associated inappropriate referrals for additional eye examination.

3.4 A pilot test of the ‘cataract filter’ algorithm

All images used in this first test were determined to be unassessable as part of a national retinal screening programme. They were initially graded as unassessable and then reviewed a second time at arbitration by an ophthalmologist as per national guidance. All retinal screening graders for Health Intelligence are trained to the same levels of competence, and take part in monthly external quality assurance exams. They also have their grading work audited every quarter to determine sensitivity and specificity. Preliminary findings from this study were reported at the IEEE Imaging Systems and Techniques conference (Russell et al., 2012), (see Appendix B).

The cataract filter was applied to previously unassessable images and the processed images were re-graded. The graders on this occasion could not use the brightness and contrast tools they are accustomed to using on the grading panel as this is not possible to do within the audit software.
Consent for testing was acquired from the patients prior to screening as no harm could come from this operation. Two parameters control the cataract filter. The "dark level" is used to mitigate the effect/extent of noise in the input image and ranges from 0 to 2. The cataract filter estimates the scattering component of light within the image as a smooth surface. The level of “smoothness” is set by the second control parameter (the scale) which ranges from 1-15%. This limits the scale over which the enhancement is allowed to adapt according to the local image properties. The higher the scale the ‘smoother’ the coverage and the less local the effect of filtering.

Processing of the retinal images was carried out in the following stages as organised by the author:

Stage 1 - 100 images previously graded as unassessable taken from the DRS system

Stage 2 - Stage one images were filtered with the cataract filter with the dark level set at 0.33 and the scale set to 5% (the default settings within the software). The processed images were then re-graded by the author to determine if any were now assessable purely as a result of the filter. Any deemed gradable would pass on to a second full disease grader (SFDG) who was not aware of the authors grade for confirmation

Stage 3 - 100 more images taken from the DRS system previously graded as unassessable (so as not to be familiar to the graders) to be prepared for a second test

Stage 4 – Adjustments were made to the filters parameters to attempt to enhance sensitivity. Firstly five images were selected from the previous 100 unassessable images and were processed through the cataract filter with the dark level set to 0, 0.5, 0.25, 0.75, 1, 1.25, 1.75 and 2 and the scale parameter set to
1,2,3,4,5,6,7,8,9,10,11,12,13,14 and 15%. This generated 600 images to review (5 images x 8 dark levels x 15 scale levels). The best settings for the cataract filter were agreed by discussion between two expert graders (the author and a medical retina ophthalmologist). These were dark range set at 1 and scale set between 4 -14 depending on the individual image.

Stage 5 – a new data set were generated with dark level set to 1 and range set to 4,5,6,7,8,9,10.11.12.13 and 14% for all 100 images which generated 1400 images to review. From this expanded dataset the settings of dark range 1 and scale 5 were determined to be the optimum settings by discussion between the two expert graders (the author and a medical retina ophthalmologist).

- Stage 6 - The 100 images taken from stage 3 were filtered with the new settings agreed in stage 5 and the images re-assessed in a standard audit fashion. Any images now deemed gradable were passed on to a second full disease grader for confirmation (this grader was an experienced second full disease grader).

3.5 Results

In stage 2, the cataract filter was used at the default settings as delivered from the manufacturer (dark level 0.33 scale 5%). Ten of the images presented after the cataract filter was applied were now deemed gradable by the graders. In order to try to refine the cataract filter settings, five unassessable images that displayed a good variation (Figure 38) were selected from the 100 in order to keep the numbers of
resulting processed images to a manageable level for detailed review. The five images
displayed varied levels of pigmentation and brightness and were selected by the
author and a medical retina ophthalmologist, low contrast and one with asteroid
hyalosis (in order to see how artefacts might be affected). As mentioned above these
five images were processed with combinations of dark level from 0.5 - 2 and scale of
1-15% generating 600 images for review, each of these images were then assessed.
The assessment of a retinal image takes approximately 2.5 mins to complete, the
grader must zoom in pan and apply various filters such as the ‘red free’ filter which enables a higher contrast to fully assess an image. The assessment took approximately 1500 min (25 man hours) to complete.

![Figure 38: The five images selected for detailed investigation. These images display a good variation in colour and features. a),b), and c) were selected as they display variation in pigmentation from asian patients. d) contains a good example of asteriod hyalosis and e) is typical of a caucasian retina.](image)
Some distortions were noted for bright reflex images when the scale parameter was set to 4% or less and so a scale parameter of 9% might be preferable to deal with the most extreme cases.

However in order to produce a simple and effective filter which can be used simple and effectively it is less desirable to have a parameter which was manually operated by the user and can open the doors to human error. It was recognised that this selection for the optimum setting was subjective, but was based on the same judgment model used by the NHS DESP and was based on a discussion between a medical retina consultant ophthalmologist and the author.
A further test took place in order to subject all 100 images to these newly refined parameters in order to determine a good average setting. As described previously all images from stage 3 were processed using 10 different sets of parameter settings (dark level 1 and scale from 4-14). Figure 39 displays an example of how varying the parameters of the cataract filter affect the assessibility of a single unassessable image.

Figure 39: An example of how varying the parameters of the cataract filter affect the quality of a single unassessable image. a) the original unassessable image. b) Image (a) after being filtered by the cataract filter with dark level 1 and scale 6%. These settings were judged to be to be too extreme by the reviewers - with some distortions noted in the darker areas. c) Image (a) after being filtered by the cataract filter with dark level 1 and scale 5% (which was judged to be the ideal setting). d) image (a) after being filtered by the cataract filter with dark
level 1 and scale 4% which was judged to be too conservative in contrast by the experts.

Typically it was felt that dark range 1 scale 4 showed too little contrast while dark range 1 scale 6 went too far and caused some distortion. This resulted in the graders choosing dark level 1 and scale 5 as it increased contrast and generated minimal distortion. When this filter was applied to the images in review 12 cases that were unassessable were now graded as assessable, of these 9 displayed clinical signs of exudate and would be classified as referable maculopathy.

![Comparison of unassessable and the 'Clearvue' filtered image. Here we see the original 'unassessable image on the left and the 'clearvue’ filtered image on the right. The circled areas not show exudate within 1 DD of the centre of the fovea and so result in a referable image.](image)

3.6 Discussion

Given that all 100 cases were originally graded unassessable (now termed inadequate by the NHS DESP) a result of 12% assessable appeared to be a significant improvement. One criticism of this pilot study is that the graders were using a system to perform the audit which differed slightly to the system which the graders use to assess images. This audit system is designed for a fast overview of cases and had
limited control over the brightness and contrast, and this may have influenced their judgment with a bias towards the filtered images. There is no current guidance on the necessity of modulating brightness and contrast in DR grading systems and in fact at the Moorfields Biomedical Research Centre for Ophthalmology graders are not allowed to use these controls as they are not currently approved by the Food and Drug Administration (FDA) in the US. The initial findings definitely look encouraging and if borne out by future work there could be substantial cost savings and benefits to patients.
Chapter 4: Conclusions and Future Work

4.1 DR HAGIS database

In Chapter 2 we detailed the development of a new database of retinal fundus images taken a DR screening programme run by Health Intelligence. This database is comprised of 39 images of DR patients many of whom suffer from comorbidities that are commonly associated with diabetic patients. These images are far more representative of current DR screening images than any freely available datasets. It is hoped that the publication of this dataset, with results from two different segmentation algorithms, may result in the development of newer and more powerful automatic segmentation tools. These 39 images span the range of assessable images that are frequently encountered in screening programs, with patients with different pigmentation, degrees of retinopathy and taken with different cameras with different resolutions. As described in chapter 2 for any automatic vessel extraction algorithm to be useful in the clinic it has to be able to deal with all these variations and complications.

4.2 Issues with the preliminary cataract filter study

As noted in chapter 3 our initial study with the cataract filter looks very promising. If this filter does in fact lead to a 12% reduction in the number of unassessable/inadequate images this would be a very useful tool. However our preliminary findings may have been influenced by:

1. The fact that the graders were using limited brightness and contrast controls in auditing compared to what they normally do when grading. The use of brightness and
contrast controls may have no effect on the results of cataract filtering, it might lead to distorted images and false alarms or it may in fact aid the graders judgement. This needs to be tested in future studies.

2. In our preliminary study the graders worked in a single sitting and were hyper-focused as they knew they were participating in a study. In the normal grading environment a screener may be less focused when grading images.

3. Experienced arbitration graders were used in order to determine the outcome of assessability. These graders are experts and will typically have a better skill set than a first full disease grader who would be the first users of this filter.

4. The images were not ‘mixed’ in with assessable images. Normally you would expect most images to be assessable again this may have biased our results.

5. It would be useful to get an automated measure of the assessibility of each test image (a measure of image quality) and then any improvement caused by the cataract filter could also be measured automatically.

### 4.3 Further cataract filter studies

It should be possible to more accurately gauge how much of an improvement had been made by employing the cataract filter if certain controls are put into place to address the issues raised in points 1-5. For example brightness and contrast tools should be made available alongside the cataract filter, as graders have become accustomed to using these in the course of their jobs.
To automatically assess the improvements made by the cataract filter an automatic retinal image quality grading algorithm (for example Hunter et al., 2011) could be implemented. This would initially allow us to see if an automatic grading algorithm was able to detect any improvement in image quality caused by the cataract filter. One such test is detailed below as Experiment A:

**Experiment A**

1. A set of images previously graded are taken from a DRS system (500) as a) adequate (assessable) or b) inadequate (unassessable).

2. The intention is to then “teach” the automatic grading algorithm (Hunter et al., 2011) how to differentiate between an assessable (adequate) and an unassessable (inadequate) image across a range of retinal screening environments. Included in these sets would be images obtained with different camera types/resolutions and from different groups.

3. Once the algorithm has determined that an image is inadequate, this can be verified by a number of graders. This will allow an analysis of the sensitivity, and specificity of the automatic grading filter.

4. Once validated the grading filter will then be used to obtain 5000 inadequate images from previously screened patient records. These 5000 images will be processed with the cataract filter and these processed images will be reprocessed by the automatic grading algorithm to see if this filter has made any difference to the automated approach.
Once an automatic algorithm to measure of image quality has been established this will allow us to compare automatic grading performance with the assessibility measures from front line retinal graders. For example:

**Experiment B**

1. The next step would be to develop two test set of images each containing 80 cases: 30 assessable unfiltered, 30 assessable filtered, 10 unassessable and unfiltered and 10 unassessable and filtered. The two sets should be counter-balanced so that the 30 assessable cases that were filtered in set 1 would be unfiltered in set 2 and visa-versa.

2. Then 20 graders sourced from East Anglia diabetic retinal screening programme (30 Graders available) would be tasked with grading either set 1 or set 2.

3. This would allow us to determine whether the cataract filter affects the perceived assessabilty of retinal images that were previously determined by retinal experts to be assessable or unassessable.

4. The two image sets could also be processed by the automatic retinal grading algorithm to see if this measure correlates with the front line grading staff.

This experiment would remove many of the concerns raised with our preliminary study. Front line graders would have access to both brightness and contrast controls for all images whether they had been filtered by the cataract filter or not. The graders used would be the first full disease graders as opposed to the experts used in our preliminary study to more accurately reflect the screening environment. All images would be presented on a Health Intelligence laptop and be similar to training systems the graders are used to. Both assessable and inassessiblle images will be presented
randomly to the graders with some but not all (50%) of each type of image having been processed by the cataract filter prior to presentation.

4.4 Conclusions

Diabetic retinopathy is currently one the leading causes of blindness in the Western world. The implementation of annual screening for DR of all diabetics over the age of 12 in the UK is reducing the number of patients becoming blind. However the global increase in the number of patients diagnosed with diabetes mellitus coupled with the increasing costs and demands on this screening system means methods to semi-automate or aid screeners are much in demand. In this thesis two preliminary studies were presented which it is hoped may aid in the development of new software tools to aid retinal graders. Chapter 2 presented a new database of DR patient fundus images with manually derived vessel masks. This database can be used to test vessel extraction algorithms and the success of two such algorithms were presented. Chapter 3 presented the application of a cataract filter to images that were deemed to be inadequate for grading by retinal screeners. Around 12% of these images changed from being inadequate to adequate through the use of this filter. Although future work is needed to confirm this finding, our preliminary results are particularly exciting.
References


Holm S, and McLoughlin N (2014). "Automatic retinal vessel extraction algorithms fro the alignment of fundus images" *PLOS One (submitted)*


Appendix A:

DR HAGIS – A Novel Fundus Image Database for the Automatic Extraction of Retinal Surface Vessels

Sven Holm, Greg Russell, Vincent Nourrit, and Niall McLoughlin

Abstract—In this paper a novel database of retinal fundus images, the DR HAGIS database, is presented and used to test the automatic extraction of the retinal surface vasculature. The DR HAGIS database consists of thirty-nine high-resolution colour fundus images obtained from a diabetic retinopathy screening programme in the UK. This NHS screening program uses a variety of services that employ different fundus cameras. This results in a range of different image sizes and resolutions depending upon where the screening takes place. In addition, patients enrolled in such screening programmes often display other co-morbidities in addition to diabetes. Therefore, in an effort to replicate the normal range of images examined by grading experts during screening, the DR HAGIS database consists not only of images of varying image resolution, but it is also divided into four co-morbidity subgroups: those with diabetic retinopathy, hypertension, age-related macular degeneration and glaucoma. Two published automatic vessel extraction algorithms were applied to this novel database. The purely intensity-based algorithm resulted in an overall mean segmentation accuracy of 95.83% (±0.67%), while the algorithm based on oriented Gabor filters had an accuracy of 95.71% (±0.66%).

Index Terms—Vessel extraction, Retina, Diabetes, Fundus image database.

I. INTRODUCTION

In the UK, eligible diabetic patients take part in a diabetic retinopathy screening programme run by the National Health Service (NHS). The aim of such a screening programme is not only to detect diabetic retinopathy (DR) but also to treat it at an appropriate stage [1]. The introduction of systematic screening has been shown to improve the cost-effectiveness of DR detection and treatment [2]. However, the number of people suffering from diabetes continues to increase, thereby increasing the workload within existing DR screening programmes [3]. It is estimated that worldwide in 2025 the number of diabetics will have increased by 122% compared to 1995, (135 million diabetics in 1995, 300 million in 2025). This trend is seen both locally and globally on a worldwide scale. In the UK the prevalence of diabetes increased by 54% between 1996 and 2005, while the incidence increased by 63% in the same period [4]. More recently, the annual report of the NHS revealed that between April 2011 and March 2012, the number of patients who were offered screening increased by 4.7%, while the number of people actually screened increased by 6.8% compared to the same period in the previous year [5].

Computer-assisted screening could help to highlight to retinal image graders images, or image regions, containing pathologies and abnormalities not easily detected otherwise [6]. As impaired oxygen supply can negatively affect the health of the retina, extracting the retinal vasculature automatically is an important tool for any computer-assisted diagnosis. For example, the automatic extraction of retinal surface vessels can be used to measure vessel diameter [7] and vessel tortuosity [8]. Changes in tortuosity has been linked to various diseases including diabetes [9] ischemic heart disease [10] and glaucoma [11]. Glaucoma has also been linked to changes in vessel diameter [12]. Moreover, the measurement of vessel diameter has been shown to provide additional predictive information on the progression of DR [13].

There is a large potential for computer-assisted diagnosis in DR screening programmes, including the use of automatic image processing algorithms, due to the increasing number of diabetics regularly screened. Before such algorithms can be implemented in a clinical setting, their accuracy needs to be verified. Several fundus image databases have been made publicly available for exactly this reason, thus allowing a comparison of the performance of various algorithms on the same dataset. Such databases exist for automatic grading of diabeticretinopathy and risk of macular edema (the MESSIDOR database [14]), detection of DR lesions (the DIARETDB1 [15] and the ROC microaneurysm set database [16]), for the calculation of retinal vessel width (the REVIEW [17] and the VICAVR database [18]), and for automatic vessel extraction (DRIVE [19], STARE [20] and ARIA [21] databases).

The largest database for vessel extraction, the ARIA database consists of one hundred and thirty-eight images taken either from healthy subjects, diabetics, or from patients with age-related macular degeneration (AMD). All images were taken with a Zeiss FF450+ fundus camera with a 50° angular field, or field-of-view (FOV). The STARE database consists of twenty fundus images, half of which were taken from healthy subjects. The images were all taken with a TopCon TRV-50 fundus camera at 35° FOV. In contrast, the forty images of the DRIVE database were all taken from diabetics with the 45° FOV setting using a Canon CR5 non-mydriatic fundus camera. Only seven of the DRIVE images show any signs of DR.

In each of these databases, the fundus images were taken with the same fundus camera in the same setting. Furthermore, the image resolutions (ARIA: 768 x 576 pixels, DRIVE: 768 x 584 pixels, STARE: 605 x 700 pixels) are significantly smaller than the image resolutions of the fundus images currently acquired in DR screening programmes across the UK. Trucio

S. Holm, G. Russell, V. Nourrit and N. McLoughlin are with the Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom e-mail: sven.holm@manchester.ac.uk

Manuscript received; revised.
et al. [6] noted that image resolution can have a large effect on the performance of vessel extraction algorithms.

The main aim of this paper is to make a more representative fundus image database, the DR HAGIS database (downloadable via http://personalpages.manchester.ac.uk/staff/niall.mcloughlin), publicly available for testing of automatic vessel extraction algorithms. This database consists of thirty-nine high-resolution images, recorded from different DR screening centres in the UK. It includes a range of different into image resolutions, and is made up of four co-morbidity subgroups, each consisting of ten images each (one patients image is duplicated into two co-morbidity subgroups). The co-morbidities included are AMD, DR, glaucoma and hypertension. In addition, two simple vessel extraction algorithms were tested against the groundtruth images provided by an expert grader. Both algorithms produced vessel extraction results compatible with an independent expert human grader.

II. MATERIALS AND METHODS

A total of thirty-nine fundus images were provided by Health Intelligence (Sandbach, UK) and were taken from diabetic patients attending a DR screening programme run by Health Intelligence. All patients gave ethical approval for the use of these images for medical research. The thirty-nine images are grouped into one of four co-morbidity subgroups: glaucoma (images 1-10), hypertension (images 11-20), DR (images 21-30) and AMD (images 31-40). One image was grouped into two subgroups, as this patient was diagnosed with both AMD and DR (images 24 and 32 are identical).

A total of three different non-mydratic fundus cameras were used to capture the fundus images: Canon CRDGi (Canon Inc, Tokyo, Japan), Topcon TRC-NW6s (Topcon Medical Systems, Oakland, NJ), and Topcon TRC-NW8 (Topcon Medical Systems, Oakland, NJ). All fundus images have a horizontal FOV of 45. Depending on the digital camera used, the images have a resolution of 4752x3168 pixels, 3565x2304 pixels, 3216x2136 pixels, 2896x1944 pixels, or 2816x1880 pixels.

Each fundus image comes with a manual segmentation of the vasculature. The surface vessels were manually segmented by an expert grader with over fifteen years experience (R.G.). These manually segmented images were taken to be the ground truth when assessing the performance of the automatic vessel extraction algorithms.

Additionally, a mask image is provided for each fundus image. The mask image depicts the area of the fundus image that contains the FOV. Only the area within the FOV is used to analyse the accuracy of the automatic vessel extraction. The mask images (M) were generated automatically. As shown in equation 1, the three channels of the fundus images (Red (R), Green (G) and Blue (B)) were added together, and a threshold value of 50 was applied to obtain a mask image. This resulted in the entire FOV being segmented as the foreground.

\[
M = (R + G + B) > 50
\]  

(1)

The two automatic vessel extraction algorithms applied to these fundus images are described in Holm and McLoughlin [22]. One algorithm is based purely on the intensity of the fundus image pixels (intensity-based) and the other is based on both the shape and intensity of the pixel patterns (Gabor based). The Gabor filter algorithm is based on the work of [23]. In short, the inverted green-channel of the RGB fundus image was used. A background estimate, obtained by applying a 100x100 pixel median kernel, was subtracted from this green-channel image. This kernel assigns the median intensity within a 100x100 pixel neighbourhood to the central pixel of this neighbourhood.

Next, pixels outside the FOV were set to the average intensity of all pixels inside the FOV to reduce any border artefacts. Twelve differently oriented Gabor filters were then applied. This resulted in an angular resolution of 15. Equation 2 shows the equation for a Gabor filter oriented at \( \theta \).

\[
g(x, y) = \frac{1}{(2\pi \sigma_x \sigma_y)} \exp \left( -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right) \cos(2\pi f x) \cos(2\pi f y) \]  

(2)

As equations 3 to 5 show, only two variables are required to define the Gabor filters: the width \( \sigma \) and the length \( \tau \) variables. These two variables determined the scale or size of the Gabor filters.

\[
\sigma_x = \frac{1}{2 \log(2) \sigma_y} \]  

(3)

\[
\sigma_y = \tau \sigma_x \]  

(4)

\[
f = \frac{1}{\tau} \]  

(5)

After applying the Gabor filters, all pixels outside the FOV were set to 0. The Gabor response images were then normalised to zero mean and unit standard deviation. A single value \( T_{Gabor} \) was applied as the threshold value to the normalised Gabor response images to generate binary masks of the vasculature. For multiscale approaches, the binary vasculature masks of each scale were summed together, and a threshold value of \( \geq 1 \) was applied. To further reduce FOV border artefacts, a mask subtraction step was included. The mask used for this step was the complement of the mask image provided with the database. To this complement image, morphological dilution with a square shaped structuring element of size 3 pixels by 3 pixels was applied, as defined in [24]. After mask subtraction, a density or bounding box filtering was applied in the final step as in n Holm and McLoughlin [22] and in Saleh and Eswaran [24]. The aim was to remove any larger objects falsely segmented as a vessel. The bounding box is defined as the smallest rectangle that can be fitted around an object. An object (i.e. a vessel segment) was removed if it had a density greater than 0.4 (see equation 6). The width, length and the threshold value \( T_{Gabor} \) were adjusted to account for the image resolution of the DR HAGIS fundus images.

\[
\text{Density} = \frac{\text{Area of object}}{\text{Area of bounding box}} \]  

(6)

\[
T = \text{MAX} + 2.5 \cdot \text{std(MAX)} \]  

(7)
The second vessel extraction method is a purely intensity-based (IB) algorithm and has been applied to the DRIVE database in Holm and McLoughlin [22]. This IB algorithm is largely based on the vessel extraction algorithm of [24]. The IB algorithm included several steps aiming at reducing the noise and illumination variation across the fundus images, thereby increasing the contrast of the vasculature. First the colour fundus images were converted into greyscale images by using the green-channel only. Since pixel intensities can be reduced when applying morphological opening, some smaller vessel segments are only segmented as vessels when morphological opening is applied, while some are only segmented when it is not applied. Therefore, the IB algorithm followed two processing pathways. One included morphological opening, while the other did not. For the morphological opening, a disk-shaped structuring element with a radius of 5 pixels was used.

All the remaining steps were common to both processing streams. As a next step, the contrast in the greyscale images (I) was enhanced using equation 8, resulting in contrast enhanced images (CE).

\[
CE = [I + TH(I)] - BH(I)
\]

The top hat (TH) and bottom hat (BH) functions enhanced both bright structures (TH) and dark structures (BH) within the fundus image. These are standard image processing functions that make use of opening (TH) or closing (BH) morphological operators. Due to the larger image resolution, the square-shaped structuring element was larger than the structuring element that was applied to the DRIVE database [22] (15x15 pixels versus 3x3 pixels, respectively).

After enhancing the contrast, the background illumination was removed. The background illumination was estimated by applying a 100x100 pixels large median kernel to image CE. The contrast enhanced image CE was subtracted from the background illumination estimate. This resulted in images with even background illumination.

In the following step, a 3x3 pixel Gaussian smoothing filter with a standard deviation of 1 pixel was applied. After Gaussian smoothing, a h-maxima transform was applied and the threshold was defined as in [24].

Each pixel was then compared to the segmentation threshold value \(T_h\) as defined in equation 7, where \(H_{MAX}\) and \(std(H_{MAX})\) are the mean and standard deviation of the h-maxima transformed image, respectively.

This resulted in two binary vessel masks, one from each processing pathway. Mask subtraction and bounding box filtering, as described for the Gabor filter approach, was applied to each vessel mask as a post-processing step. However, a density threshold of 0.3 was used here. In the final step, the two post-processed binary vessel masks were combined into single binary vessel mask. This was achieved by summing the two post-processed vessel maps together and applying the threshold value \(\geq 1\).

The quality of segmentation was determined by the mean percentage of correctly segmented pixels within the FOV. This mean accuracy is defined in equation 9, where TP, TN, FP and FN are the number of true positives, true negatives, false positive and false negative pixels, respectively. The FOV was defined by the provided mask images. Sensitivity (equation 10) was defined as the percentage of vessel pixels within the FOV segmented as such, and the specificity (equation 11) as the percentage of correctly segmented background pixels (again within FOV). Results are given in mean (± standard deviation).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

(9)

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

(10)

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

(11)

Several different values for the model variables were tested for their effect on the overall segmentation accuracy. The search space for these variables is discussed in Appendix A for the IB algorithm and in Appendix B for the Gabor filter algorithm.

III. RESULTS

Figure 1 shows typical fundus images taken from the DR HAGIS database, one from each of the four subgroups. The corresponding segmentation results are shown in figure 2 for the IB algorithm and figure 3 for the Gabor filter algorithm. The accuracy, sensitivity and specificity for all 39 fundus images are listed in table I for the IB and for the two-scale Gabor filter algorithm.

In figures 2 and 3, green pixels highlight the correctly segmented vessel pixels (true positives), red pixels the falsely segmented background pixels (false negatives), and blue pixels the falsely segmented vessel pixels, or oversegmented pixels (false positives). Black pixels within the FOV correspond to correctly segmented background pixels (true negatives).
The overall mean segmentation accuracy for the IB algorithm was 95.83% (±0.67%) with a sensitivity of 55.83% (±6.42%) and a specificity of 98.91% (±0.35%). If the segmentation threshold value used to segment the h-maxima transformed images was chosen to be the same as in [22], the sensitivity increased to 83.47% (±4.59%). However, this decreased the overall segmentation accuracy and specificity to 92.71% (±2.25%) and 93.43% (±2.51%), respectively.

For the Gabor-filter algorithm, a single-scale, a two-scale or a three-scale approach was implemented. Each scale was defined by its width and the length variables, as well as the threshold value $T_{\text{Gabor}}$. For the single-scale approach, the best results were generated when $t = 25$, $l = 0.9$ and $T_{\text{Gabor}} = 2.1$. For the two-scale approach $t_1 = 20$, $l_1 = 1.3$, $l_2 = 30$, $T_{\text{Gabor}} = 2.5$. In the three-scale approach, the Gabor filter parameters and threshold value that generated the best results were $t_1 = 15$, $l_1 = 1.7$, $l_2 = 25$, $l_3 = 0.9$, $T_{\text{Gabor}} = 2.9$.

Figure 3 shows the segmentation results for the fundus images in figure 1 using the two-scale Gabor filter approach. Across all thirty-nine fundus images, the overall segmentation accuracy for the single-scale approach was 95.68% (±0.64%), for the two-scale approach 95.71% (±0.66%), and for the three-scale approach 95.69% (±0.70%). The sensitivity was 60.10% (±6.71%), 59.68% (±7.38%), and 58.28% (±8.16%), respectively. The specificity varied from 98.43% (±0.74%) for the single-scale, to 98.50% (±0.71%) for the two-scale and 98.59% (±0.68%) for the three-scale Gabor filter algorithm.

Table II lists the segmentation accuracy, sensitivity and specificity for each of the four co-morbidity subgroups separately. For the IB algorithm, all three measures of segmentation quality were highest in the hypertensive subgroup (accuracy: 96.24% (±0.32%), sensitivity: 57.78% (±8.07%), specificity: 98.97% (±0.45%)). For the glaucoma subgroup similar results were obtained for accuracy and sensitivity (96.07% (±0.46%) and 57.44% (±6.53%), respectively. The mean specificity was identical (98.97% (±0.32%)). The segmentation quality was slightly worse for the DR and AMD subgroups (accuracy: 95.42% (±0.61%) and 95.54% (±0.52%), sensitivity: 54.38% (±5.08%) and 53.01% (±5.26%), specificity: 98.72% (±0.35%) and 98.90% (±0.30%), respectively.

Similar results were obtained for the three different Gabor filter approaches. The single-, two- and three-scale approaches resulted in similar accuracies, sensitivities and specificities across all four co-morbidity subgroups. However, the highest accuracies were obtained for the glaucoma subgroup (96.10% (±0.39%), 96.15% (±0.45%) and 96.14% (±0.62%) for the single-, two- and three-scale, respectively), followed by the hypertension subgroup (95.94% (±0.52%), 95.91% (±0.60%), 95.86% (±0.71%), respectively). Similar to the IB algorithm, the vessel extraction for the AMD subgroup was slightly more accurate than for the DR subgroup when using any of the three Gabor filter approaches (AMD: 95.31% (±0.69%), 95.36% (±0.72%), 95.33% (±0.76%) and DR: 95.16% (±0.68%), 95.23% (±0.68%), 95.23% (±0.70%), respectively). The Gabor filter approaches extracted a larger proportion of the retinal vasculature, resulting in higher sensitivity across all four subgroups. The highest sensitivity was obtained for the glaucoma subgroup (62.90% (±5.52%), 62.65% (±6.26%), 61.13% (±8.88%), for the single-, two- and three-scale, respectively), and lowest for the AMD subgroup (56.97% (±7.60%), 56.54% (±8.02%), 55.25% (±8.42%), respectively). However, as shown in table II, the specificity was slightly lower for the Gabor filter approaches than for the IB algorithm. The highest specificity was obtained for the glaucoma subgroup (98.82% (±0.49%), 98.70% (±0.48%), 98.80% (±0.62%), respectively). The DR subgroup showed the lowest specificity (98.02% (±0.96%), 98.12% (±0.92%),...
98.20% (±0.88%), respectively.

### IV. DISCUSSION AND CONCLUSION

Of the three implementations of the Gabor filter algorithm, the two-scale implementation resulted in the highest accuracy. However, all three implementations (single-, two- and three-scale) resulted in very similar overall mean segmentation accuracy, sensitivity and specificity. Likewise, both the IB algorithm and the Gabor filter approaches resulted in a similar overall segmentation accuracy (IB algorithm: 95.83%, two-scale Gabor filter algorithm: 95.71%). For the DRIVE database, the accuracy of the second observer was 94.73% [19]. If a similar inter-observer accuracy is assumed for the DR HAGIS database, we can conclude that both the IB algorithm and Gabor filter approach perform as well as an expert human observer in terms of overall segmentation accuracy.

The DR HAGIS database consists of four co-morbidity subgroups. These subgroups highlight the typical lesions, both pathological and due to laser photocoagulation, seen in a normal DR screening programme. Such lesions can make it difficult to automatically extract the retinal surface vasculature without oversegmenting the images. This trade-off between a high sensitivity and a high specificity likely explains the higher segmentation threshold values used here compared to
the DRIVE database implementation in Holm and McLoughlin [22]. Furthermore, this would explain the lower sensitivity of both vessel extraction approaches implemented here. However, the use of oriented Gabor filters, which detect elongated structures similar to blood vessels, did improve the sensitivity. It is possible that the vessel extraction algorithms presented here could be improved by optimising the segmentation parameters for either each of the four co-morbidity subgroups or for each image resolution, a task that, however, is beyond the scope of this study.

The tortuosity and diameter of retinal vessels have been shown to change in response to both ocular and cardiovascular diseases [9]–[12]. Furthermore, impairment in ocular blood flow has been implicated in disease states such as diabetes [25]–[28], glaucoma [29], [30] and other ocular diseases [31]. It is crucial for any automatic analysis of the retinal blood vessels that the extraction of these vessels is both possible and accurate enough, in particular if the automatic analysis of the vasculature provides clinicians with additional diagnostic information. Therefore, we have put together a novel retinal fundus image database consisting of typical fundus images taken from a DR screening programme and made it publicly available to allow others to test their automatic image processing software. The images contain different co-morbidities, image resolutions, and were taken using different fundus and digital cameras to reflect the variability in the datasets encountered in the clinic. If an automatic vessel extraction algorithm is to be used effectively in a clinical setting, it must tackle all these challenges.

**TABLE II**

<table>
<thead>
<tr>
<th>Test type</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DR algorithm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>95.42%</td>
<td>54.38%</td>
<td>98.72%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>96.24%</td>
<td>57.78%</td>
<td>98.97%</td>
</tr>
<tr>
<td>AMD</td>
<td>95.54%</td>
<td>53.01%</td>
<td>98.90%</td>
</tr>
<tr>
<td>Glaucma</td>
<td>96.07%</td>
<td>57.44%</td>
<td>98.97%</td>
</tr>
<tr>
<td><strong>Single-scale Gabor filter algorithm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>95.16%</td>
<td>59.96%</td>
<td>98.02%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>95.94%</td>
<td>60.16%</td>
<td>98.53%</td>
</tr>
<tr>
<td>AMD</td>
<td>95.31%</td>
<td>56.97%</td>
<td>98.34%</td>
</tr>
<tr>
<td>Glaucma</td>
<td>96.10%</td>
<td>62.90%</td>
<td>98.62%</td>
</tr>
<tr>
<td><strong>Two-scale Gabor filter algorithm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>95.23%</td>
<td>59.72%</td>
<td>98.12%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>95.91%</td>
<td>59.25%</td>
<td>98.57%</td>
</tr>
<tr>
<td>AMD</td>
<td>95.36%</td>
<td>56.54%</td>
<td>98.42%</td>
</tr>
<tr>
<td>Glaucma</td>
<td>96.15%</td>
<td>62.65%</td>
<td>98.70%</td>
</tr>
<tr>
<td><strong>Three-scale Gabor filter algorithm</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>95.23%</td>
<td>58.68%</td>
<td>98.20%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>95.86%</td>
<td>57.45%</td>
<td>98.66%</td>
</tr>
<tr>
<td>AMD</td>
<td>95.33%</td>
<td>55.25%</td>
<td>98.50%</td>
</tr>
<tr>
<td>Glaucma</td>
<td>96.14%</td>
<td>61.13%</td>
<td>98.80%</td>
</tr>
</tbody>
</table>

# Appendix A

**Search Space for the IB Algorithm**

Several different values were tested for most of the variables to refine the IB algorithm for vessel extraction. In a first round of refinement, the threshold value used to segment the fundus images (the 2.5 in equation 7) was varied between 0.5 and 1.5 in intervals of 0.1, and between 1.7 and 2.9 in intervals of 0.2. Similarly, the size of the median filter used to obtain an estimate of the background illumination was set to either 80x80 pixels or to 100x100 pixels. The threshold value for the small object removal step was varied between 240 - 640 pixels in steps of 100 pixels. The density threshold was varied from 0.1 to 0.9 in intervals of 0.4.

After this first round of refinement, the segmentation threshold value was set to 2.5. However, the size of the median filter was set to either 80×80 pixels or to 90×90 pixels. The small object threshold value was varied from 160 - 340 pixels in intervals of 40 pixels, while the density threshold value was varied from 0.3 to 0.7 in intervals of 0.2.

In a third round of refinement, the median filter was set to either 90×90 pixels, 100×100 pixels, or 110×100 pixels. The small object removal threshold value was set between 80 - 160 pixels in intervals of 40 pixels. And the density threshold value was either 0.1 or 0.3.

The small object removal threshold value was varied between 60 - 120 pixels in intervals of 20 pixels in a fourth round of refinement. The density threshold value was set to 0.1, 0.3 or 0.5, and the size of the median filter varied as in the third round of refinement.

After this fourth round of refinement, ten different Gaussian smoothing filters were tested to find the most effective Gaussian filter. Not applying a Gaussian smoothing filter did not improve the segmentation accuracy (data not shown). Furthermore, the radius for the circular structuring element used in the morphological opening step was set to either 1, 5 or 25, and the small object removal threshold value was varied between 0 - 80 pixels in intervals of 20 pixels. After refining all the variables, the effect of the different post-processing steps
on the overall segmentation accuracy was studied to develop the final IB algorithm discussed above. All other variables were defined as in Holm and McLoughlin [22].

**Appendix B**

**Search Space for the Gabor Filter Algorithm**

Several different values for the segmentation threshold value and for width and length factors of the Gabor filters (and I, respectively) were tested to measure the accuracy of the Gabor filter algorithm. For the single-, two and three-scale implementation, the segmentation threshold value was varied from 1.3 to 4.0 in intervals of 0.4. For the single-scale implementation, the width factor was varied between 5 and 45 (in intervals of 5), and the length factor was varied from 0.9 to 3.3 in intervals of 0.4. For the two-scale approach, the width factor was varied between 5 and 25 (interval: 5) for the smaller Gabor filter and between 25 and 45 (interval: 5) for the larger Gabor filter. For both Gabor filters, the same length factors were used as in the single-scale implementation. For the three-scale implementation, the width factor for the small Gabor filters were set to 5, 10 or 15, the medium-sized Gabor filters were set to 20, 25 or 30, and the large Gabor filters were set 35, 40 or 45. For each of the three Gabor filters the same length factors were tested as for the single- and two-scale implementations.

The size of the median filter used to generate the background illumination estimates was set to 100x100 pixels, as in the optimised IB algorithm. All other variables were kept as in Holm and McLoughlin [22].

**Acknowledgment**

The authors would like to thank Health Intelligence for kindly providing the fundus images that make up this novel DR HAGIS database.

**References**


Appendix B:

Enhancement of color retinal images in poor imaging conditions

G. Russell,
Health Intelligence Ltd. Sandbach, UK
Greg.russell@health-intelligence.com

J. P. Oakley
University of Manchester, Faculty of Engineering and Physical Sciences, Manchester, UK
john.oakley@manchester.ac.uk

N. McLoughlin, V. Nourrit
University of Manchester, Faculty of Life Sciences Manchester, UK
vincent.nourrit@manchester.ac.uk

Keywords-component: retinal imaging, scattering.

I. INTRODUCTION

One of the first accounts of retinal observation in vivo is perhaps Jean Mery’s now famous report to the French Royal Academy of Sciences in 1684. The French anatomist described how immersing a cat underwater allowed him to observe the retina in more details [1]. Nowadays a large range of retinal imaging instruments exists (e.g. flood imaging, optical scanning laser ophthalmoscopy, optical coherence tomography, etc. [2]) together sophisticated techniques to correct for the eye’s optical aberrations (e.g. adaptive optics, deconvolution [3]).

Aberrations are however only one of the two main optical phenomena degrading the retinal image’s quality. As light propagates to and from the retina, it will be scattered by small inhomogeneities within the ocular media [4]. The amount of scattering in a young non pathological eye is low but will increase with age and some pathologies (e.g. cataract). This increased scattering will result in a degradation of the retinal image in terms of reduced brightness, poorer contrast, colour degradation and possibly lower spatial resolution.

Such a problem is similar to degradation of images by the atmosphere. Although a large body of work exist on the enhancement of images recorded in poor visibility conditions, a paucity of papers have focused on reducing the degradation of retinal images by intraocular scattering [5].

In this context, our aim was to adapt and assess the potential of an automatic defogging system to retinal imaging.

II. METHODS

A. Descatering

Image contrast enhancement can be divided into two main types: model based and non model based. Model based algorithms improve image contrast by reversing the underlying cause of image degradation but requires a large amount of information about the geometry of the scene and the nature of the scattering. Due to the difficulty to develop a realistic yet tractable model of intraocular scattering, we took interest in non-model based algorithms as they do not require information about the cause of degradation.

Perhaps the most important non-model based algorithm is histogram equalization (HE). This algorithm is usually performed on the luminance and saturation of the image but not the hue so as to maintain the original color. However, because intraocular scattering is partly wavelength dependent [6], this method may cause some color degradation.

Oakley and Bu [7] presented a successful non-model based algorithm to restore the chrominance and luminance of a scene while maintaining good color fidelity. In this model, the image (I) recorded by the camera is the sum of the light reflected by the scene (here the retina (R)) as it would be observed in the absence of scattering together with the scattered light, known in the remote sensing literature as “airlight” (A).

\[ I = R + A \] (1)

The airlight value will depend on the distribution and the concentration of the different scattering particles (e.g. keratocytes in the cornea, proteins aggregates in the crystalline lens), the distance from the eye to the camera and the angle of illumination. As a result, the airlight value can vary across the image. Another version of this algorithm has recently been reported that can deal with such spatial variation [8]. The degradation model assumed in both [7] and [8] is an additive noise with a possible scaling of the R term due to extinction. In extremely turbid conditions this model breaks down and significant blurring takes place due to small-angle scattering and recovery of a clear image becomes difficult. However in previous work it was found that the additive noise is the most significant image defect in atmospheric imaging. It is therefore reasonable to assume that the same may be the case in retinal imaging but it is obviously important to check this empirically.

B. Artificial images
We first assessed the potential of the technique on retinal images artificially degraded. The images were obtained with an artificial eye and a standard fundus camera. After acquisition an airlight value was added to the images to simulate various level of scattering. This allowed us to assess the optimum value for the adjustable parameters (one that controls the level of spatial variation of the airlight estimate and one that controls the noise compensation) and to quantify the ability of the algorithm to estimate the airlight value.

C. Real images

The industrial partner, Health Intelligence Ltd, provided us with a series of retinal images obtained in patients with various degrees of ocular opacities and recorded with a similar camera.

III. RESULTS

![Figure 1](image1.jpg)  
Figure 1. Row 1, image not degraded with corrected version on the right. Row 2, image artificially degraded with corrected version on the right. Row 3, image artificially degraded (strongly) with corrected version on the right.

Results obtained with artificial images are illustrated in figure (1). Row 1 shows an image of an artificial eye, with the enhanced image to the right. Rows 2 and 3 show the same image but with the contrast deliberately reduced to 30% and 15% using Matlab software with the following command lines:

\[
\text{im2} = \text{im1} \times 0.30 + 0.70 + 0.004 \times \text{randn(size(im1))};
\]

\[
\text{im3} = \text{im1} \times 0.15 + 0.85 + 0.004 \times \text{randn(size(im1))};
\]

where im2 and im3 are output images, im1 is the input image (scaled to unity with no gamma encoding) and the ‘randn’ function generates a Normally-distributed random number that represents a realistic noise level.

As can be observed, the “correction” does not significantly degrade the image if no scattering is present (i.e. no airlight).

When the visibility is very poor, e.g. in case of advanced cataract, the recovered images show a grainy effect. This is due to the scaling needed to restore image contrast after subtraction of the airlight [8]. However, in spite of this defect, the recovered images show a good level of consistency when the input contrast falls to low levels such that the retinal vessels become almost completely invisible.

Results obtained with real images are illustrated in figure (2) and these show a similar effect. The first row presents the original images. On the second row, the images are corrected by histogram equalization. This processing was performed using Photoshop CS5.1 from Adobe Corporation. On the third row, images are improved using a more sophisticated contrast enhancement technique (i.e. Contrast-limited adaptive histogram equalization (CLAHE) [9]). Finally, images corrected using our technique are presented on the bottom row.

Image quality is by nature subjective, particularly for images of low to moderate quality as the ones processed, and although few metrics have been suggested for retinal images, there is not yet an accepted gold standard [10]. In addition, the quality of an image measured with one of these metrics may not correlate well with the suitability of an image for diagnosis purposes. For these reasons, the processed images presented in Figure 2 were assessed subjectively by an expert retinal grader.
As can be observed simple histogram equalization provides little improvement and significantly alter the color balance. Results presented on the last 2 rows appear more beneficial. The color balance in images processed with CLAHE appears however strongly altered and less natural which reduces the possibility to differentiate between various clinical elements such as arteries and veins, or drusens and exudates. Important artefacts are also present as bands on the right and left sides of the images. If less dramatic, such artefacts could be misinterpreted as pathological feature (e.g. choroidal fold).

IV. DISCUSSION

A novel method to improve the quality of retinal image degraded by scattering is presented. The main advantage is the ability to improve automatically the contrast and color fidelity of the image. These improvements were superior to those achieved with locally adaptive histogram equalization.

Such improvements could be highly beneficial in the context of screening for Diabetic Retinopathy (DR). Diabetic patients regularly undergo fundus photography and if the image is considered “inadequate” by retinal graders the patient will usually be referred to additional examination leading to more stress for the patient, possibly a delayed diagnosis and increased costs for the health care system.

Current methods used in retinal screening to enhance images rely basically on simple brightness and contrast manipulations which can easily lead to loss of information (e.g. blown out highlights) or the creation of spurious clinical information (e.g. colour distortion may lead a pigment to look like a microaneurysm or fine vessels to be confused with a microvascular abnormality). Our method overcome these faults of colour and feature distortion, allowing the colour levels and contrast to be optimised automatically to offer the grader additional information. This should help reducing the rate of ungradable images (current for most screening programmes is higher than the 7% recommended rate by the DESP) and associated inappropriate referrals for additional eye examination.

We plan now to test our algorithms on a large database of images to confirm that it will not convey any erroneous clinical information (e.g. to hide an existing lesion or to create a non-existing one) and quantify the results.

REFERENCES


