

Tomi Kauppi

EYE FUNDUS IMAGE ANALYSIS FOR AUTOMATIC DETECTION OF DIABETIC RETINOPATHY

Thesis for the degree of Doctor of Science (Technology) to be presented with due permission for public examination and criticism in the Auditorium of the Student Union House at Lappeenranta University of Technology, Lappeenranta, Finland on the 12th of December, 2010, at noon.

Acta Universitatis
Lappeenrantaensis 414

- Supervisor Professor Heikki Kälviäinen
Professor Joni-Kristian Kämäräinen
Adjunct Professor Lasse Lensu
- Machine Vision and Pattern Recognition Laboratory
Department of Information Technology
Faculty of Technology Management
Lappeenranta University of Technology
Finland
- Reviewers Associate Professor Alfredo Ruggeri
Laboratory of Biomedical Imaging
Department of Information Engineering
University of Padova
Italy
- D. Sc (Tech) Jussi Tohka
Academy of Finland research fellow
Department of Signal Processing
Tampere University of Technology
Finland
- Opponents Professor Ela Claridge
School of Computer Science
The University of Birmingham
United Kingdom

ISBN 978-952-265-016-0
ISBN 978-952-265-017-7 (PDF)
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Digipaino 2010



To Pirjo and our beloved baby.

Preface

The work presented in this thesis has been carried out during the years 2006-2010 as part of the IMAGERET-project. The project was a joint effort of the Machine Vision and Pattern Recognition Laboratory in University of Lappeenranta, Department of Ophthalmology in University of Kuopio and Color Research Laboratory in University of Joensuu. Several people have directly or indirectly supported the work and it is pleasure show my gratitude to all those who made this thesis possible.

First of all I would like to thank my supervisors Professor Heikki Kälviäinen for the practical and financial supervision, Professor Joni Kämäräinen and Adjunct Professor Lasse Lensu for showing the meaning of true science. It has been a privilege to have the benefit of your counsel.

I would like to express my gratitude to co-authors Professor Hannu Uusitalo from the University of Tampere, Professor Iiris Sorri, Valentina Kalesnykiene, Asta Raninen and Raija Voutilainen from the University of Eastern Finland, and Juhani Pietilä from the Perimetria Oy. I would also like to thank all the other people involved in the IMAGERET-project and especially Professor Jussi Parkkinen, Dr. Markku Hauta-Kasari, Dr. Jouni Hiltunen and Pauli Fält from the University of Eastern Finland, and Professor Majid Mirmehdi from the University of Bristol.

I am very grateful for the reviewers Associate Professor Alfredo Ruggeri and especially Dr. Jussi Tohka for the valuable comments that clearly made the thesis better. Also, I would like to thank Professor Ela Claridge who promised to be my opponent.

For the financial support I would like to thank the Finnish Funding Agency for Technology and Innovation (TEKES), Kuomed Oy, Mawell Oy, Perimetria Oy and Santen Oy. Also, the support from the Lappeenranta University of Technology research foundation (Lauri and Lahja Hotinen foundation) is greatly appreciated.

For inspiring working environment I would like to thank my co-workers, Arto, Ilmari, Jani, Janne, Jarkko, Jarmo, Jukka, Jussi, Leena, Olli, Pekka, Saku, Teemu T., Teemu K., Toni, Ville, and all the rest. Special thanks go to Tuomas Eerola for the peer support, numerous fruitful discussions and pool games (which I won) over the last six years.

The words cannot truly express the gratitude that I owe to my parents, sister and especially my loving family Pirjo and Venla.

Lappeenranta, November 2010

Tommi Kauppi

Abstract

Tomi Kauppi

Eye Fundus Image Analysis for Automatic Detection of Diabetic Retinopathy

Lappeenranta, 2010

150 p.

Acta Universitatis Lappeenrantaensis 414

Diss. Lappeenranta University of Technology

ISBN 978-952-265-016-0, ISBN 978-952-265-017-7 (PDF), ISSN 1456-4491

Diabetes is a rapidly increasing worldwide problem which is characterised by defective metabolism of glucose that causes long-term dysfunction and failure of various organs. The most common complication of diabetes is diabetic retinopathy (DR), which is one of the primary causes of blindness and visual impairment in adults. The rapid increase of diabetes pushes the limits of the current DR screening capabilities for which the digital imaging of the eye fundus (retinal imaging), and automatic or semi-automatic image analysis algorithms provide a potential solution.

In this work, the use of colour in the detection of diabetic retinopathy is statistically studied using a supervised algorithm based on one-class classification and Gaussian mixture model estimation. The presented algorithm distinguishes a certain diabetic lesion type from all other possible objects in eye fundus images by only estimating the probability density function of that certain lesion type. For the training and ground truth estimation, the algorithm combines manual annotations of several experts for which the best practices were experimentally selected. By assessing the algorithm's performance while conducting experiments with the colour space selection, both illuminance and colour correction, and background class information, the use of colour in the detection of diabetic retinopathy was quantitatively evaluated.

Another contribution of this work is the benchmarking framework for eye fundus image analysis algorithms needed for the development of the automatic DR detection algorithms. The benchmarking framework provides guidelines on how to construct a benchmarking database that comprises true patient images, ground truth, and an evaluation protocol. The evaluation is based on the standard receiver operating characteristics analysis and it follows the medical practice in the decision making providing protocols for image- and pixel-based evaluations. During the work, two public medical image databases with ground truth were published: DIARETDB0 and DIARETDB1. The framework, DR databases and the final algorithm, are made public in the web to set the baseline results for automatic detection of diabetic retinopathy.

Although deviating from the general context of the thesis, a simple and effective optic disc localisation method is presented. The optic disc localisation is discussed, since normal eye fundus structures are fundamental in the characterisation of DR.

Keywords: Diabetic retinopathy detection, eye fundus imaging, benchmarking image database, eye fundus image processing, eye fundus image analysis, optic disc localisation, medical image processing, medical image analysis

UDC 004.932.2 : 611.84

SYMBOLS AND ABBREVIATIONS

$\alpha_0, \alpha_1 \dots, \alpha_n$	Parameters of radial polynomial illuminance model
α_c	Weight of c th Gaussian mixture model component
$\beta_0, \beta_1, \dots, \beta_n$	Parameters of bivariate polynomial illuminance model
ϵ	Sum of squared distances between reference landmark set and aligned training landmarks sets
ϵ_1, ϵ_2	Imaging errors; ϵ_1 : shot and thermal noise, and ϵ_2 : quantisation error, amplifier noise, D/A and A/D noise
ζ	Diagnostic test produced score values for test subjects
ζ^{im}, ζ^{pix}	Image analysis algorithm produced image and pixel score values for test images
$\zeta_{\xi_i}^{im}, \zeta_{\xi_i}^{pix}$	Baseline algorithm produced image and pixel score value for lesion type ξ_i
θ_{ξ_i}	Parameter list of Gaussian mixture model probability density function for lesion type ξ_i
λ	Wavelength of light
$\lambda(x, y)$	Customised similarity metric between template and image patch at image point (x, y)
μ_c	Mean of c th Gaussian mixture model component
ν	Pixel intensity value
$\hat{\nu}$	Histogram matched intensity value
ξ	Set containing DIARETDB1 lesions types $\{HA, MA, HE, SE\}$
ξ_i	i th lesion type in set ξ
$\bar{\xi}_i$	Background class for the lesion type ξ_i
σ	Smoothing factor for kernel density estimation
σ^s, σ^t	Standard deviations for given $l\alpha\beta$ colour space component in source and target image (colour transfer)
τ	Rotation of observed planar object around the axis parallel to the optical axis in Kang-Weiss illuminance model
$\phi(x, y)$	Generic notation for parametric illuminance model value at point (x, y)
$\phi_c(x, y)$	Cosine fourth law of illumination model value at point (x, y)
$\phi_e(x, y)$	Elliptic paraboloid illuminance model value at point (x, y)
$\phi_k(x, y)$	Kang-Weiss illuminance model value at point (x, y)
$\phi_p(x, y)$	Bivariate polynomial illuminance model value at point (x, y)
$\phi_r(u, v)$	Radial polynomial illuminance model value at point (x, y)
$\varphi(x, y)$	Non-uniform image illuminance factor for pixel (x, y)
χ	Rotation of observed planar object around the x -axis in Kang-Weiss illuminance model
ω	True clinical states for test subjects
$\omega^{im}, \omega^{pix}$	True clinical states for test images and test image pixels
Σ_c	Covariance of c th Gaussian mixture model component
$\Psi_{X,E}$	Colour decorrelated template space

A_k	Off-axis illumination term in Kang-Weiss illuminance model
a, b	Lower and upper integration limits for the partial area under the curve
$B(x, y)$	Blue component pixel value in RGB image
$B_0(x, y)$	Blue component pixel value in illuminance corrected RGB image
C	Component count in Gaussian mixture model
c^s	Source image pixel value for given component in $l\alpha\beta$ colour space (colour transfer)
\hat{c}	Colour transferred source image pixel value for given component in $l\alpha\beta$ colour space (colour transfer)
$\langle c^s \rangle, \langle c^t \rangle$	Source and target image means for given component in $l\alpha\beta$ colour space (colour transfer)
$C(\lambda)$	Light reflected from the retina
C_{FNR}	Cost of false negative rate
C_{FPR}	Cost of false positive rate
$CC(x, y)$	Cross-correlation between a template and an image patch at image point (x, y)
$d^2(x, y)$	Squared Euclidean distance between a template and an image patch at image point (x, y)
D_i	Binary map representing manual annotations of i th medical expert
$D_{\xi_i}^j$	Confidence map representing manual annotations of j th medical expert for lesion type ξ_i
$D_i(x, y)$	Binary decision of i th medical expert that an object is present at image point (x, y)
$D_{\xi_i}^j(x, y)$	Subjective confidence of j th medical expert that an object is present at image point (x, y)
e_1, e_2, e_3	Eigenvectors that span colour decorrelated template space
E	Minimum overall difference between ground truth \mathbf{GT} and all the expert annotations D_i
\mathbf{E}	Matrix containing eigenvectors that span colour decorrelated template space
f	Focal length
$f_c(\cdot)$	Radiometric response function of the camera
$f(t)$	Function to divide given t into two domains to prevent an infinite slope at $t = 0$ in $L^*a^*b^*$ colour space
$\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n$	Image features extracted from eye fundus image
$g(i, j)$	Image patch value at point (i, j)
\hat{G}_k	Radial falloff term in Kang-Weiss illuminance model
$G(x, y)$	Green component pixel value in RGB image
$G_0(x, y)$	Green component pixel value in illuminance corrected RGB image
\mathbf{GT}	Ground truth estimated from multiple expert annotated image locations
$GT(x, y)$	Binary pixel value of the ground truth estimated from multiple expert annotated image locations
h_{ij}	Element of the matrix that defines a linear transformation between two colour spaces

i, j	Generic enumeration variables
$I(x, y)$	Intensity value at image point (x, y)
$I_0(x, y)$	Intensity value at illuminance corrected image point (x, y)
K	Number of training landmark sets
L	Ordered sequence of unique pixel score values in ζ^{pix}
L_0	Irradiance at the image centre
$L_{sampled}$	Unique pixel score values sampled from L
$L(\lambda)$	Light source spectrum
$l_{\xi_i}(\mathbf{x})$	Likelihood ratio between the lesion type ξ_i and the corresponding background class
$\mathcal{L}(\cdot)$	Likelihood function for simultaneous ground truth and expert annotation performance estimation
l, α, β	Components in $l\alpha\beta$ colour space
\hat{n}_{ξ_i}	Number of experts that have made annotations in the image for the lesion type ξ_i
n, m	Generic variable defining a count or number
N	Set containing $X\%$ of the largest $p(\mathbf{x} \xi_i)$ in the image (summax decision rule)
N_a	Correction plane in additive image illuminance correction
$NCC(x, y)$	Normalised cross-correlation between a template and an image patch at image point (x, y)
$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$	Multivariate normal distribution of a colour pixel value \mathbf{x} , where $\boldsymbol{\mu}_c$ is the mean vector and $\boldsymbol{\Sigma}_c$ is the covariance matrix
$\mathcal{N}(\nu; \nu_i, \sigma^2)$	Univariate normal distribution of an intensity value ν , where ν_i is the mean and σ^2 is the variance
N_x, M_y	Width and height of an optic disc template
$MOP_{\xi_i}(x, y)$	Mean confidence for the lesion type ξ_i to be present in the pixel location (x, y) according to \hat{n} experts
\mathbf{O}	Optic disc colour pixel values in the training set
$\bar{\mathbf{O}}$	Mean optic disc colour in the training set
$p_1, p_2 \dots p_9$	Parameters of the elliptic paraboloid illuminance model
$p(\zeta, normal)$	Probability density function of a normal test population with respect to the diagnostic test produced score value
$p(\zeta, abnormal)$	Probability density function of an abnormal test population with respect to the diagnostic test produced score value
$p(\mathbf{x} \xi_i)$	Multivariate Gaussian mixture model probability density function of the lesion type ξ_i with respect to a colour pixel value \mathbf{x}
$p(\mathbf{x} \bar{\xi}_i)$	Multivariate Gaussian mixture model probability density function of the backgroundclass $\bar{\xi}_i$ with respect to a colour pixel value \mathbf{x}
$p(\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n)$	Output of the eye fundus image analysis algorithm for the eye fundus image features $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n$
$p_s(\nu), p_r(\nu)$	Observed and reference image probability density functions with respect to an intensity value ν (histogram matching)
r	Distance from image centre
\hat{R}	Cost ratio between the false negative rate and the false positive rate

$r(x, y)$	Light received by the pixel coordinate (x, y)
$R(x, y)$	Red component pixel value in RGB image
$R_0(x, y)$	Red component pixel value in illuminance corrected RGB image
$\hat{r}, \hat{g}, \hat{b}$	Normalised RGB values, where $r + g + b = 1$
r_i, g_i, b_i	Generic colour coordinates in trichromatic colour space
$R(\lambda), G(\lambda), B(\lambda)$	Spectral sensitivities of the camera sensors
$\overline{R}, \overline{G}, \overline{B}$	Components of an arbitrary colour space \overline{RBG} constructed from RGB colour space by using a linear transformation
R_y, R_x, R_z	Rotation terms of the elliptic paraboloid model around the indicated coordinate axis
$s(x, y)$	Similarity map value at point (x, y)
$S(\lambda)$	Reflective properties of the retina
SN_i, SP_i	Sensitivity and specificity pair denoting annotation performance of i th medical expert in STAPLE ground truth estimation algorithm
SN, SP	Annotation performances of n medical experts
\hat{SN}, \hat{SP}	Estimated annotation performances of n medical experts
$t_{\xi_i}^{pix}$	Decision threshold that lesion type ξ_i is present in the pixel
$t_{\xi_i}^{img}$	Decision threshold that lesion type ξ_i is present in the image
thr_{vote}	Voting threshold
$t(i, j)$	Template value at point (i, j)
$T_i(\cdot)$	Linear transformation for i th landmark set
T_k	Object tilt term in Kang-Weiss illuminance model
$T_r(\nu), T_s(s)$	Cumulative density functions of reference and observed image with respect to intensity value ν (histogram matching)
u, v	Spatial image location (origin at the centre of the image)
\mathbf{x}	Colour pixel value (e.g. RGB)
$\overline{\mathbf{X}}$	Reference landmark set of n landmarks
\mathbf{X}	Landmark set of n landmarks
$X\%$	Percentage notation for summax decision rule
x, y	Spatial image location (origin in the upper left image corner)
$\overline{x}, \overline{y}$	Rotated and translated pixel coordinates of elliptic paraboloid model
X_N, Y_N, Z_N	Reference white point coordinates in XYZ colour space
\mathbf{Y}	Matrix containing vectorised image pixel values
z	Value of translated and scaled elliptic paraboloid model
3-D	Three dimensional
AF-FM decomposition	Amplitude modulation - frequency modulation decomposition
AFP	Average number of false positives
AUC	Area under the ROC curve
BRISTOLDB	Non-public eye fundus image database collected in University of Bristol
CAD	Computer-aided detection or diagnosis

CALTECH101	Image Dataset for object categorisation
CCD	Charged-coupled device
CIE L*a*b*	Perceptually uniform colour space: lightness (L*), red-green (a*), yellow-blue (b*))
CIE XYZ	Red (X), Green (Y) and Blue (Z) colour space for standard observer
CMIF	Collection of multispectral images of the fundus
DET curve	Detection error trade-off curve
DIARETDB0	Diabetic retinopathy image database 0
DIARETDB1	Diabetic retinopathy image database 1
DR	Diabetic retinopathy
DRIVE	Digital retinal images for vessel extraction
DTD	Document Type Definition
EER	Equal error rate
FCM	Fuzzy C-means clustering
FN	Number of false negatives
FNR	False negative rate
FP	Number of false positives
FPR	False positive rate
FROC	Free-response receiver operating characteristics curve
GMM	Gaussian mixture model
GMM-FJ	GMM using Figuerdo-Jain parameter estimation
GUI	Graphical user interface
HA	Haemorrhage(s)
HE	Hard exudate(s)
HL7	Health Level Seven International
HSV	Hue (H), saturation(S) and value (V) colour space
IRMA	Intraretinal microvascular abnormalities
KNN	K-nearest-neighbour (classification)
LABELME	Open image annotation tool
LMS	Colour space represented by the responses of the three cones type in the human eye: long (L), medium (M) and short (S)
LUV	Perceptually uniform colour spaces: lightness (L) and chromaticities axes (U and V)
MA	Microaneurysm(s)
NCC	Normalised cross-correlation
NPV	Negative predictive value
MESSIDOR	Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology
PAUC	Partial area under the ROC curve
PCA	Principal component analysis
PNG	Portable network graphics
NPV	Negative predictive value
PDF	Probability density function
PPV	Positive predictive value
REVIEW	Retinal vessel image set for estimation of widths

RGB	Red (R), green (G) and blue (B) colour space
ROC curve/analysis	Receiver operating characteristics curve/analysis
ROC database	Retinopathy online challenge database
SE	Soft exudate(s) also known as cotton-wool spots
SIFT	Scale-invariant feature transform
SN	Sensitivity
SP	Specificity
STARE	Structured analysis of the retina
STAPLE	Simultaneous truth and performance level estimation
SVD	Singular value decomposition
SVM	Support vector machine
SYM	Symmetry point
TESD	Truth estimation from self distances
TN	Number of true negatives
TNR	True negative rate
TP	Number of true positives
TPR	True positive rate
WER	Weighted error rate
XML	Extensible markup language

1	Introduction	19
1.1	Background	19
1.2	Research questions	20
1.3	Restrictions	20
1.4	Contributions	20
1.5	Structure of the thesis	21
2	Eye and diabetes	23
2.1	Structure and function of the eye	23
2.2	Diabetic eye diseases	25
2.2.1	Diabetic retinopathy	26
2.2.2	Cataract	28
2.2.3	Neovascular glaucoma	30
2.2.4	Diabetic neuropathies	30
2.3	Diagnosing diabetic retinopathy	30
2.4	Screening diabetic retinopathy	35
2.5	Automatic detection of diabetic retinopathy	36
2.6	Summary	46
3	Benchmarking of eye fundus image analysis algorithms	47
3.1	Introduction	47
3.2	Requirements for benchmarking	49
3.2.1	True patient images	49
3.2.2	Ground truth given by experts	50
3.2.3	Evaluation protocol	50
3.3	Eye disease databases	50
3.4	Patient images and ground truth	52
3.4.1	Collecting patient images	52
3.4.2	Image annotations as the ground truth	54
3.4.3	Data format for medical annotations	55
3.5	Algorithm evaluation	56
3.5.1	Evaluation methodology	56
3.5.2	ROC-based quality measures	59
3.5.3	Image-based evaluation	62
3.5.4	Pixel-based evaluation	63
3.6	Baseline algorithm	64
3.7	Case studies – DIARETDB databases	66
3.7.1	DIARETDB0	67
3.7.2	DIARETDB1	67
3.7.3	BRISTOLDB	70
3.8	Summary	71
4	Overall image score and multiple expert information	73
4.1	Introduction	73

4.2	Overall image score selection	74
4.3	Ground truth estimation from multiple expert annotation	75
4.4	Classifier training from multiple expert information	80
4.5	Experimental results and discussion	80
4.6	Conclusion	86
5	Photometric cue in lesion detection	87
5.1	Introduction	87
5.2	Related work	90
5.3	Trichromatic photometric information	90
5.4	Colour space selection	91
5.4.1	CIE L*a*b*	91
5.4.2	HSV	92
5.4.3	Experimental results and discussion	93
5.5	Utilising background class information	93
5.5.1	Experimental results and discussion	94
5.6	Image illuminance correction	95
5.6.1	Imaging and illuminance	96
5.6.2	Illuminance estimation and its correction	96
5.6.3	Parametric illuminance models	97
5.6.4	Experimental results and discussion	99
5.7	Colour correction	101
5.7.1	Histogram matching	101
5.7.2	Colour transfer	103
5.7.3	Geometric colour correction using image landmarks	104
5.7.4	Experimental results and discussion	106
5.8	Applicability of the image analysis results	107
5.8.1	Results and discussion	108
5.9	Under- and overexposed pixels in eye fundus images	110
5.9.1	Results and discussion	110
5.10	Conclusion	114
6	Optic Disc Localisation	117
6.1	Introduction	117
6.2	Related work	118
6.3	Pre-processing	118
6.4	Optic disc extraction	120
6.4.1	Colour decorrelated template space	120
6.4.2	Template matching	121
6.5	Experiments	124
6.6	Discussion	126
6.7	Conclusion	126
7	Discussion	128
	Bibliography	133
	Appendix A Database characteristics	

Appendix B Image-based evaluation results (ROC curves)

Appendix C Pixel-based evaluation results (ROC curves)

Appendix D Training error of baseline algorithm

Appendix E Example images of DIARETDB1

Appendix F Illuminance corrected example images

Appendix G Colour corrected example images

Appendix H Illuminance and colour corrected example images

1.1 Background

Diabetes, which can be characterised as a chronic increase of glucose in the blood, has become one of the most rapidly increasing health threats worldwide [190, 191]. There are an estimated 150 to 200 million people diagnosed with diabetes, of which approximately 50 million within Europe alone [23]. Moreover, a large number of people remain undiagnosed. In Finland, which has a population of around 5 million, there are 280 000 people under diabetes care of which insulin production in the pancreas is permanently damaged for 40,000 people (type 1 diabetes), and resistance to insulin is increased for 240,000 people (type 2 diabetes) [189]. In addition, the current estimates predict that there exist 200,000 undiagnosed patients and that the number of people receiving diabetes care will double every 12 years. These alarming facts promote prevention strategies and screening over a large population since proper and early treatment of diabetes is cost-effective [159].

Digital imaging technology has developed into a versatile non-invasive measurement tool which enables a wealth of applications also in medical sciences. Imaging the eye fundus with modern techniques is a current practise in many eye clinics, and it is becoming even more important as the expected lifetime and the costs of health care increase. Since the retina is vulnerable to microvascular changes of diabetes and diabetic retinopathy is the most common complication of diabetes, eye fundus imaging is considered a non-invasive and painless route to screen and monitor such diabetic eyes [174].

Since diagnostic procedures require attention of an ophthalmologist, as well as regular monitoring of the disease, the workload and shortage of personnel will eventually exceed the current screening capabilities. To cope with these challenges, digital imaging of the eye fundus, and automatic or semi-automatic image analysis algorithms based on image processing and computer vision techniques provide a great potential [11, 127]. By automating the analysis process, more patients can be screened and referred for further examinations, and the ophthalmologists have more time for patients that require their attention since most of the eye fundus images are not leading to any medical action.

1.2 Research questions

Diabetic retinopathy is diagnosed from an eye fundus image and the grading is based on identifying lesions, i.e., morbid changes in colour, texture or shape in tissue or organs. An essential cue to decide whether an eye fundus image contains such lesions is the photometric information, that is, the information resulting from the measurement of light. In the present automatic eye fundus image analysis, however, the photometric cue is often overlooked and the effort is mainly put on other cues, such as shape and texture which motivates the first research questions

1. How to utilise the photometric cue in the detection of diabetic retinopathy, and
2. How good is the photometric cue in the detection of diabetic retinopathy?

To study these research questions, and to enable a fair comparison with the existing approaches, a public eye fundus image database with verified ground truth and solid performance evaluation protocol is required. In this way, the performance can be evaluated and the choices made justified. In addition, the performance of any other method in literature can be evaluated and compared. This motivates the final research question:

3. How the performance of the diabetic retinopathy detection algorithms should be evaluated to produce reliable comparison?

The main objective of this thesis is to answer the research questions.

1.3 Restrictions

The research was limited to persons with clearly visible symptoms of diabetic retinopathy and persons with no diabetic abnormalities in the eye fundus. The following signs of diabetic retinopathy were studied: microaneurysms, haemorrhages and exudates (hard and soft).

1.4 Contributions

One main contribution of the thesis is the framework and public databases for benchmarking eye fundus image analysis algorithms. During the course of the work, two medical image databases with ground truth were published: DIARETDB0 and DIARETDB1. The DIARETDB0 database was published as a comprehensive research report [87], and the DIARETDB1 database was originally reported in [86, 85].

While collecting the eye fundus image databases, DIARETDB0 and DIARETDB1, it became evident that collecting benchmarking databases from the scratch is demanding, laborious and time consuming, and therefore the practical issues and frequently occurring sub-tasks are discussed in this thesis.

Another important contribution is how to utilise photometric cue in the detection of diabetic retinopathy. The role of image illuminance correction in eye fundus images was

published in [90], and the research related to the most important aspects in the use of photometric information, such as colour space selection, learning and classification of colour cues, and both image illuminance and colour correction, are discussed in this thesis.

In addition, two problems essential for supervised learning and classification in eye fundus image processing was addressed in [89]: 1) how to fuse medical annotations collected from several medical experts into unambiguous ground truth and for training a classifier, and 2) how to form an image-wise overall score for accurate and reliable automatic eye fundus image analysis.

During the research visit in the University of Bristol, the optic disc localisation was studied as a supplementary item. The research results were reported in [88].

The author made the major contribution to the development, writing and experimental work in [87, 86, 85, 88, 89, 90].

1.5 Structure of the thesis

Chapter 2 contains the physiological background concerning the structure and function of the eye. It is followed by the description of diabetes related eye diseases and their symptoms which are important for the current diagnostic procedures. Finally, the current and future prospects of the early detection of diabetic retinopathy are discussed including the potential and current state of the automated diagnosis.

Chapter 3 provides guidelines on how to construct benchmarking databases for eye fundus image analysis algorithms. The guidelines describe on how to collect patient images, ground truth, and how to use them in performance evaluation. The given results and tools are utilised to establish a benchmarking database, DIARETDB1, for detection of diabetic retinopathy.

Chapter 4 discusses the two problems essential for the supervised learning and classification in eye fundus image processing: 1) how to fuse medical annotations collected from several medical experts into unambiguous ground truth and for training a classifier, and 2) how to form an image-wise overall score for accurate and reliable eye fundus image analysis. As a conclusion, a baseline algorithm for DIARETDB1 is devised.

Chapter 5 investigates the use of colour in detection of diabetic retinopathy. The chapter covers the colour space selection, use of background class information, and both image illuminance and colour corrections. Moreover, the applicability of the results, and the effect of under and overexposed pixels are discussed. Conclusions are given based on the experimental results.

Chapter 6 discusses the optic disc localisation in colour eye fundus images and describes a simple and robust optic disc localisation method.

Chapter 7 summarises the achievements and proposes future directions.

In this chapter, the diabetic complications in the eye and their implications to vision are discussed. The chapter contains the physiological background concerning the structure and function of the eye, and the description of diabetes-related eye diseases and their symptoms. For the most common diabetic eye disease, diabetic retinopathy, the diagnostic procedures and modalities are presented, and the current and future prospects of early detection are discussed. The discussion includes the shortcomings of the current diagnosis and the potential benefits of automated eye fundus image analysis.

2.1 Structure and function of the eye

In the optical sciences, the human eye is often compared to a camera [177]. Light reflected from an object is focused on the retina after passing through the cornea, pupil and lens, which is similar to light passing through the camera optics to the film or a sensor. In the retina, the incoming information is received by the photoreceptor cells dedicated for detecting light. From the retina, the information is further transmitted to the brain via the optic nerve, where the sensation of sight is produced. During the transmission, the information is processed in the retinal layers. A cross-section of the eye and the structures involved in the image formation are presented in Fig. 2.1.

There are three important features in the camera which can be seen analogous to the function of the eye: aperture, camera lens, and the camera sensor. In the eye behind the transparent cornea, the coloured iris regulates the amount of light entering the eye by changing the size of the pupil [68]. In the dark, the pupil is large allowing the maximum amount of light to enter, and in the bright the pupil is small preventing the eye to receive an excess amount of light. In the same way, the camera regulates the amount of light entering the camera with the aperture. In order of the eye to focus on objects at different distances, the ciliary muscle reshape the elastic lens through the zonular fibres. For objects in short distances, the ciliary muscle contracts, zonular fibres loosen, and the lens thickens into orb shaped which results high refractive power. When the ciliary muscle is relaxed, the zonular fibres stretch the lens into thin shaped and the distant

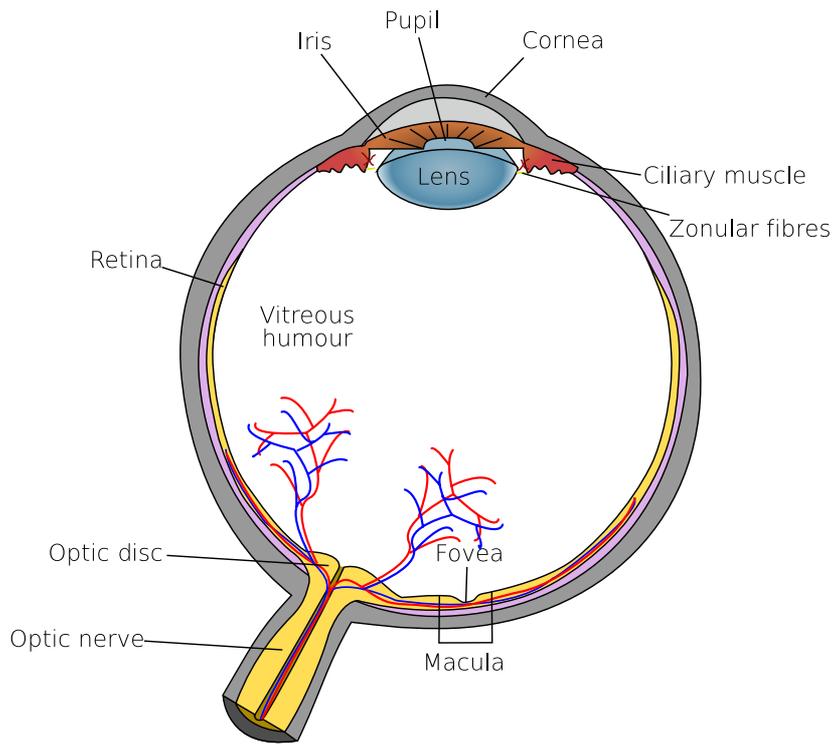


Figure 2.1: Cross-section of the eye (modified from [185]).

objects are in focus. This corresponds to the function of focal length, i.e. the distance between the lens and sensor, when focusing the camera. If the eye is properly focused, the light passes through the vitreous gel to the camera sensor of the eye, that is the retina.

The retina is the inner surface of the eye and consists of transparent tissue of several layers of cells designated to absorb and convert the light into neural signals [68]. The order of the retinal layers is peculiar since the conversion is carried out by the light detecting photoreceptor cells on the layer which is in the back of the retina and furthest from the light. Thus, the light has to travel through the retinal layers before it reaches the photoreceptor cells [162]. Once the light is detected, converted and the neural signals collected to the optic nerve, the impulses are finally transmitted to the brain. During transmission from the photoreceptor cells to the optic nerve the electric impulses are further processed in the inner layers of the retina.

The detailed central vision is formed in the macula which is a highly light sensitive area 5 to 6 mm in diameter in the central region of the retina [49]. In the centre of the macula is a round shaped area known as fovea, where the cones are almost exclusively found. The cones are photoreceptor cells selectively sensitive to different wavelengths of light. Next to the macula is the beginning of optic nerve (optic nerve head or optic disc), from where the main artery and vein emerge in the retina. There are no normal retinal layers

in this region and therefore the absence of photoreceptor cells results in a blind spot in the retina. The nutritional support to the retina is provided by the choroid and the two main capillary networks: the nerve fibre layer network and the connecting neuron layer network [162]. The capillary density increases towards the centre region of the retina and the most dense network is found in the macula, but the fovea itself is absent of capillaries. Therefore, the fovea is dependent on the choroidal blood supply from the vascular layer behind the retina (choroid). The presented anatomical parts (macula, fovea, capillaries, and optic nerve head) highlighted in Fig. 2.2 are the relevant structures of the retina in terms of retinal diseases and this thesis.

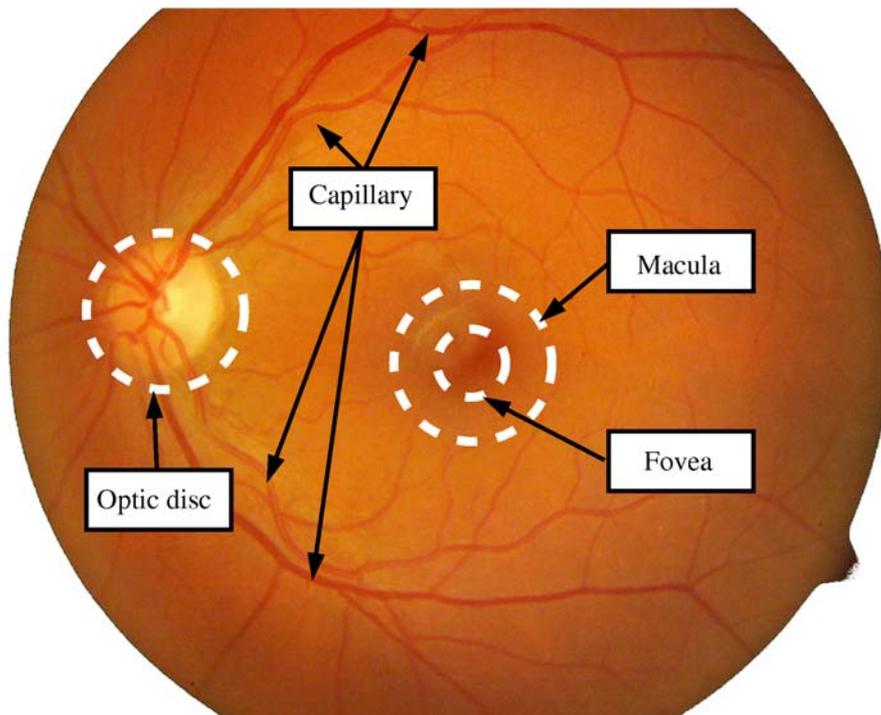


Figure 2.2: Normal physiological parts of the eye fundus.

2.2 Diabetic eye diseases

There are a number of reasons that can cause reduced visual acuity, visual impairment, and blindness. In diabetic eye diseases, the cause of visual disturbances is in most cases related to those vascular changes diabetes is causing to the eye. The discussion in this section concentrates on the diabetic eye diseases that encompass a group of eye problems, such as diabetic retinopathy, cataract, neovascular glaucoma and diabetic neuropathies [110]. The section discusses how the symptoms of the diabetic eye diseases emerge and how they affect the vision. The effect of the diabetic eye diseases on vision is illustrated in Fig. 2.3.



Figure 2.3: Influence of diabetes on vision: (a) normal vision; (b) diabetic retinopathy; (c) cataract; (d) neovascular glaucoma (Courtesy: National Eye Institute, National Institutes of Health [110]).

2.2.1 Diabetic retinopathy

Diabetic retinopathy is a microvascular complication of diabetes, causing abnormalities in the retina. Typically there are no salient symptoms in the early stages, but the number and severity predominantly increase in time. In the following, the progress of the disease is described in detail.

The diabetic retinopathy typically begins as small changes in the retinal capillaries. The smallest detectable abnormalities, microaneurysms (MA), appear as small red dots in the retina and are local distensions of the weakened retinal capillary (Fig. 2.4(a)). Due to these damaged capillary walls, the small blood vessels may rupture and cause intraretinal haemorrhages (HA). In the retina, the haemorrhages appear either as small red dots indistinguishable from microaneurysms or larger round-shaped blots with irregular outline

(Fig. 2.4(b)). The diabetic retinopathy also increase the permeability of the capillary walls which results in retinal oedema and hard exudates (HE). The hard exudates are lipid formations leaking from the weakened blood vessels and appear yellowish with well-defined borders (Fig. 2.4(c)). If the local capillary circulation and oxygen support fail due to obstructed blood vessels, pale areas with indistinct margins appear in the retina. These areas are small microinfarcts known as soft exudates (Se) (Fig. 2.4(d)). Intraretinal microvascular abnormalities (IRMA) and venopathy are signs of a more severe stage of diabetic retinopathy, where intraretinal microvascular abnormalities appear as dilation in the capillary system and venopathy as shape changes in artery and veins. An extensive lack of oxygen and obstructed capillary in the retina lead to the development of new fragile vessels. These new vessels attempt to grow towards the suffering tissue to supply nutrition and oxygen. However, the new vessels are fragile and tend to grow into the space between the retina and vitreous humour, or directly to the vitreous humour, which can lead to preretinal haemorrhage and a sudden loss of vision. The growth of these new vessels is called neovascularisation. (Fig. 2.4(e)).

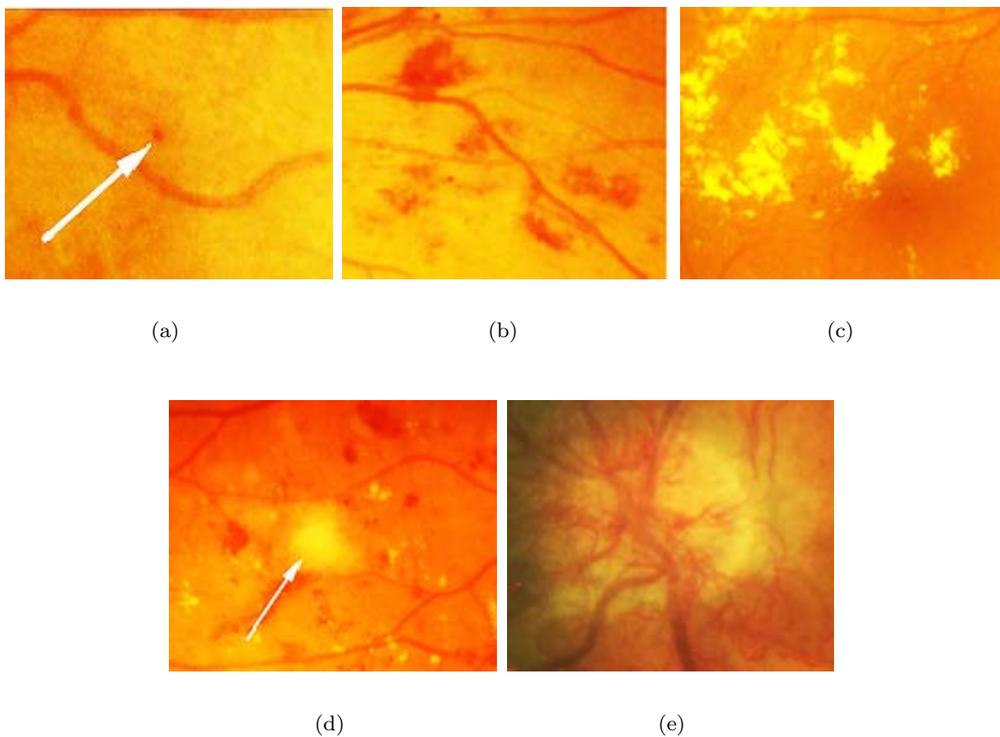


Figure 2.4: Symptoms of diabetic retinopathy (images processed for better visualisation): (a) microaneurysm; (b) haemorrhages; (c) hard exudates; (d) soft exudate; (e) neovascularisation in optic nerve head.

STAGES OF DIABETIC RETINOPATHY AND MACULOPATHY

The severity of diabetic retinopathy is divided into two stages: nonproliferative (background retinopathy) and proliferative retinopathy. The nonproliferative retinopathy indicates the presence of diabetic retinopathy in the eye and consist of microaneurysms, haemorrhages, exudates, retinal oedema, IRMA and venopathy [186, 29]. The microaneurysms and especially hard exudates typically appear in the central vision region (macula) which predicts the presence of macular swelling (macular oedema). The symptoms of nonproliferative retinopathy and the macular swelling characterise the maculopathy which is the most common cause of visual disability among the diabetic people. [186, 29]. Although, the maculopathy may occur at any stage of the diabetic retinopathy, it is more likely in the advanced stages of the disease. In the worst case, it can result irreversible damage to the fovea [162]. A retina with nonproliferative retinopathy is illustrated in Fig. 2.5 and a retina with maculopathy is illustrated in Fig. 2.6.

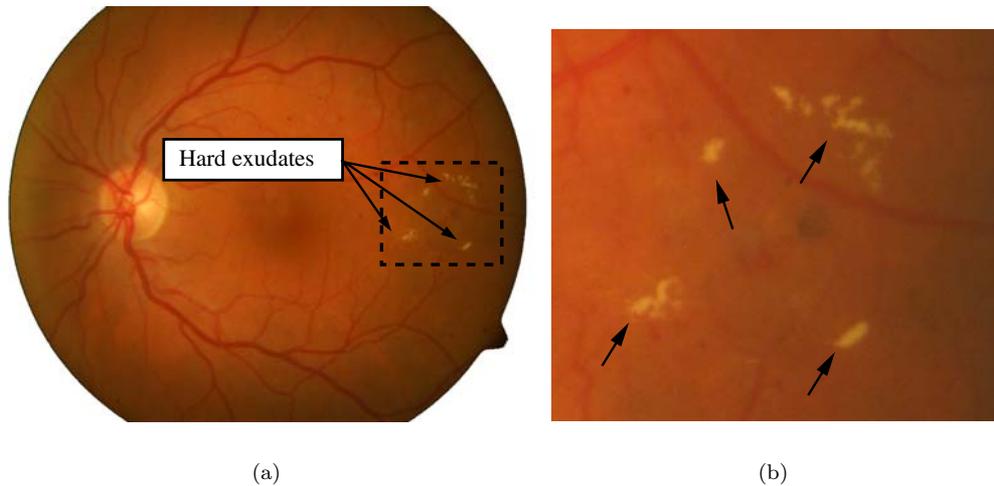


Figure 2.5: Example of nonproliferative diabetic retinopathy: (a) eye fundus image showing hard exudates; (b) close up image of the hard exudates.

If the nonproliferative retinopathy is untreated or undiagnosed it will turn into proliferative retinopathy which is also an eye-sight threatening condition. The proliferative diabetic retinopathy may cause sudden loss in visual acuity or even a permanent blindness due to vitreous haemorrhage or tractional detachment of the central retina. This stage is considered if neovascularisation or vitreous/preretinal haemorrhage is present in the retina [186, 29]. A retina with proliferative retinopathy is illustrated in Fig. 2.7.

2.2.2 Cataract

Cataract is defined as a decrease in the clarity of the lens which gradually degrades the visual quality [110]. In hyperglycaemia, the opacification in the posterior pole of the

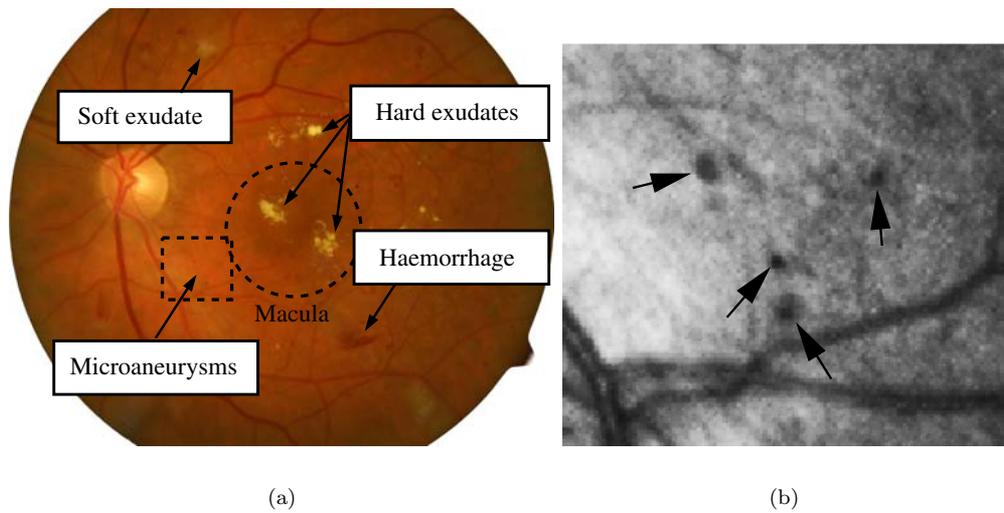


Figure 2.6: Example of maculopathy: (a) eye fundus images with maculopathy showing haemorrhages, microaneurysms, exudates (soft and hard); (b) close-up image of microaneurysms.

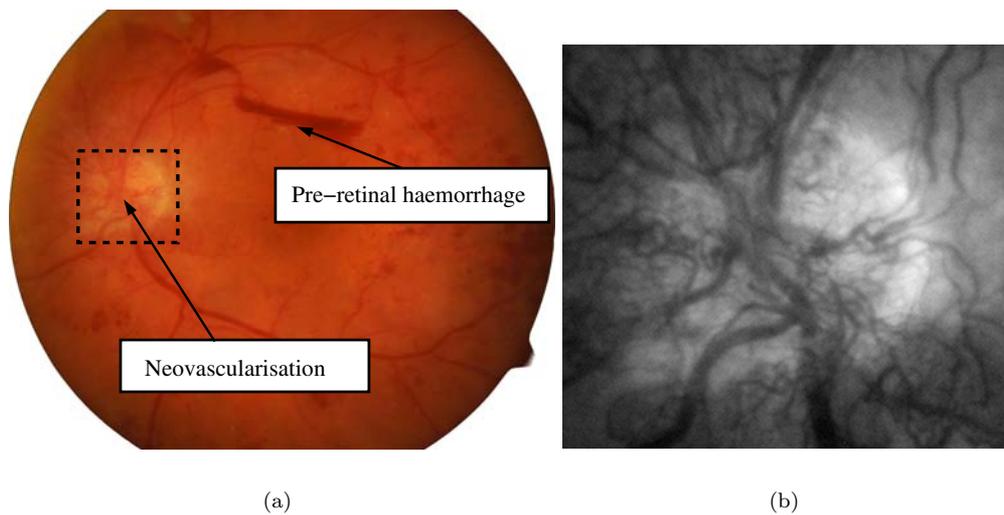


Figure 2.7: Example of proliferative diabetic retinopathy: (a) eye fundus image showing pre-retinal haemorrhage and neovascularisation; (b) close up image of neovascularisation in the optic nerve head (zoomed from contrast stretched green channel).

lens is caused by the changed metabolism of the lens epithelial cell (posterior subcapsular cataract). Since the lens is responsible for focusing light to the retina, the cataract blocks and distorts the light passing through the lens making the imaging of eye fundus difficult. Therefore, a cataract is a common annoyance in the diagnosis of diabetic retinopathy. Typical visual effects are decreased sensitivity to the light, blurred vision, difficulty with glare and dulled colours (Fig. 2.3(c)). The disease is common for older people since it is usually related to aging and develops gradually in time. In rare occasions, the disease is present at birth or in early childhood, but there are several reasons for the disease to occur earlier in life, such as severe eye trauma, uveitis and diabetes. It is approximated that cataract occur 10-15 years earlier in people with diabetes which is related to the fluctuation of the blood sugar levels [162].

2.2.3 Neovascular glaucoma

The failure of microcirculation in the eye can cause the growth of new vessels in the iris and the chamber angle resulting acute raise in intraocular pressure [162, 153]. This condition is neovascular glaucoma which may occur in people with diabetes due to the ischemic nature of the proliferative retinopathy. The neovascular glaucoma may develop without symptoms, but many experience pain, red eye, light sensitivity or decreased vision. However, the growth of new vessels in the iris and the chamber angle is considered as highly advanced stage of diabetic eye disease which is difficult to cure and often results serious vision loss and therefore the early treatment is essential. The effect of neovascular glaucoma on vision is illustrated in Fig. 2.3(d).

2.2.4 Diabetic neuropathies

Diabetes can also temporarily affect the optic nerve and nerves controlling the eye movement such as nervus oculomotorius (III), trochlearis (IV) and abducens (VI) [162]. The diabetic neuropathies typically cause temporary cross-eyedness that can be alarming for the patient, but it does not indicate permanent damage.

2.3 Diagnosing diabetic retinopathy

Diabetic retinopathy is the most common complication of diabetes and the primary cause for visual impairment and blindness in adults. In this section, the diagnosis of diabetic retinopathy is discussed and the main diagnostic modalities are briefly described.

The diagnosis of diabetic retinopathy is based on clinical eye examination and eye fundus photography [79]. The self diagnosis of diabetic retinopathy is extremely difficult if diabetes is not suspected, verified from the blood samples or visual impairment is not present. Thus, making diabetic retinopathy a devious eye disease. The eye fundus photography is the preferred diagnostic modality for the primary health care, and for the cases where retinal fundus photographs are ungradeable or unavailable, the clinical eye examination is used. Alternate modalities [104], such as fluorescein angiography and optical coherence tomography, are typically utilised to reinforce the eye examination. If the retina is unreachable and light cannot traverse in the eye, the condition of the retina can be inspected using ophthalmic ultrasound. However, the ultrasound cannot

directly detect diabetic retinopathy, but it can detect if retinal detachment is present due to proliferative retinopathy. It is important to note that it is not possible to diagnose diabetic retinopathy using laboratory tests.

In the screening of diabetic retinopathy, the primary health care doctor use either retinal photography (the first eye fundus photograph evaluation) or direct ophthalmoscopy to investigate the state of the retina [29]. Patients having either no or mild changes are monitored in the primary health care. If the symptoms are in the more advanced stage or the eye fundus images are ungradeable, the patient is referred to an ophthalmologist, preferably specialised in diabetic retinopathy. The ophthalmologist re-evaluate or take new eye fundus images (the second eye fundus photograph evaluation), or conduct clinical examinations to diagnose the severity of the disease. Depending on the diagnosis, the patients are appointed for further examinations or treatment. A flowchart of diagnostic procedures is illustrated in Fig. 2.8.

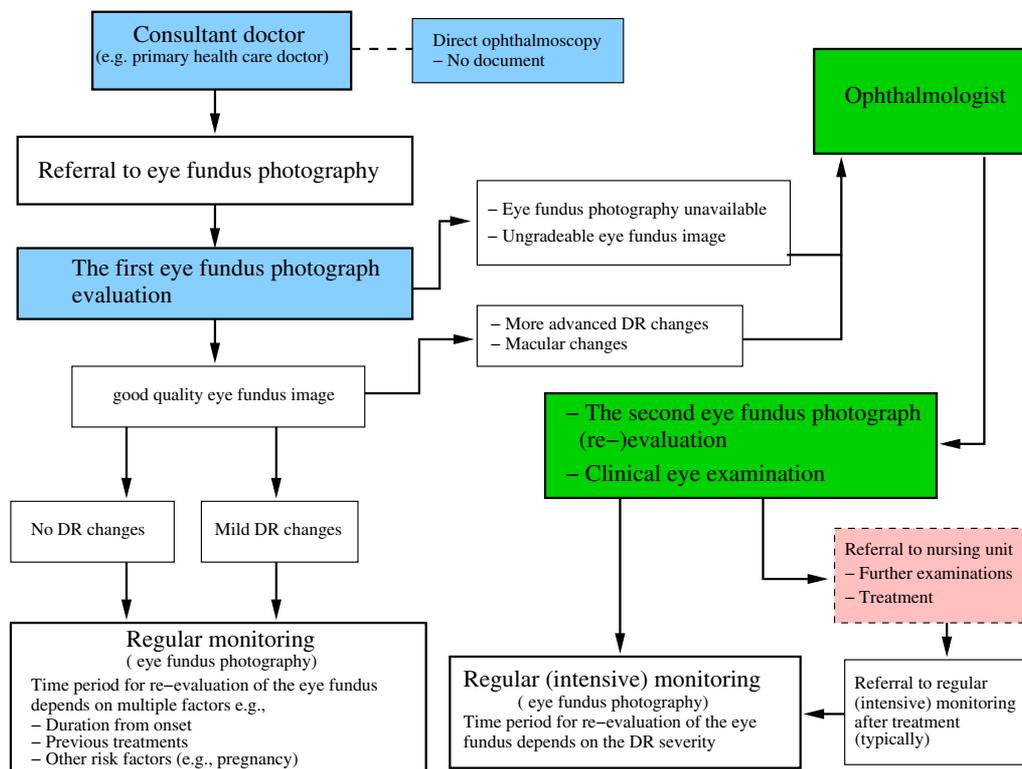


Figure 2.8: A flowchart for diagnosing and monitoring diabetic retinopathy. Blue and green filled boxes with solid borders denote the phases in the eye examination and the responsible medical expert, and the red filled box with the dashed border the treatment phase (modified from [29]).

CLINICAL EYE EXAMINATION

Main tools in clinical eye examination are direct and indirect ophthalmoscopes, and biomicroscope with indirect lenses. A direct ophthalmoscope is a hand held apparatus through which a medical expert can observe the patient's eye. The apparatus consists of the illumination source and corrective lenses, where the light beams are reflected into to the patient's eye using a mirror or prism [68]. In the indirect ophthalmoscopy, the patient's eye is examined from an arm's length by focusing high intensity light through a hand-held condensing lens to the patient's eye and examining the reflected light (stereoscopic image) with the binocular lenses. The illumination source and the binocular lenses are mounted in a medical expert worn headband. The biomicroscope comprises an observation system and illumination system, where the observation system is a biomicroscope capable of wide range of magnifications and the illumination system emits focal light into the patient's eye that can be controlled with slit mechanism and apertures [65]. Combined with wide field retinal lenses, large areas of the retina can be visualised.

EYE FUNDUS PHOTOGRAPHY

As mentioned, eye fundus photography is considered the preferred diagnostic modality if available since it is reliable, non-invasive and easy to use [29]. In contrast to traditional ophthalmoscopy, it allows to record diagnostic data and enable the expert consultation afterwards, and more importantly the eye fundus photography results in a better sensitivity rate, that is, a better detection rate of abnormal eye funduses [79]. Due to the rapid development of digital imaging, the eye fundus cameras also provide easy to file images in portable format that enable automatic diagnosis of diabetic retinopathy using image analysis algorithms. An eye fundus camera is illustrated in Fig. 2.9.

Eye fundus cameras are divided into two groups: mydriatic and non-mydriatic cameras, where the prefix denotes the requirement for dilation of the pupils with eye drops. The prefix is misleading since in practice the dilation is employed in both fundus camera types. Non-mydriatic fundus cameras are smaller and suitable for screening purposes, but at the same time the image quality is worse and the field-of-view smaller. Thus, mydriatic cameras are used when a more accurate diagnosis is needed.

The patient is seated in front of the fundus camera and the head is positioned into the instrument's head rest. A flash lamp produced light is emitted into patient's eye using optical mirrors and lenses and the reflected light is captured by the camera sensor. The captured images are typically colour images (Fig. 2.10(a)), but since the retina is transparent and the penetration depth of the emitted light depends on the wavelength, the desired retinal structures can be emphasised using optical filters. A typical alternative for colour images for diagnosing diabetic retinopathy are the red-free eye fundus images (Fig. 2.10(b)). The recommendation in the case of diabetic retinopathy diagnosis is to use the both red-free and colour images. [176], where two images are captured by focusing the 45° field-of-view fundus camera to macula and optic disc (two-field 45° fundus photography) [14]. For long-term diabetic patients, two-field 60° photography is recommended since the neovascular changes that need treatment are typically found in the periphery, even if the changes in the central areas of the retina are slight [175].



Figure 2.9: An eye fundus camera.

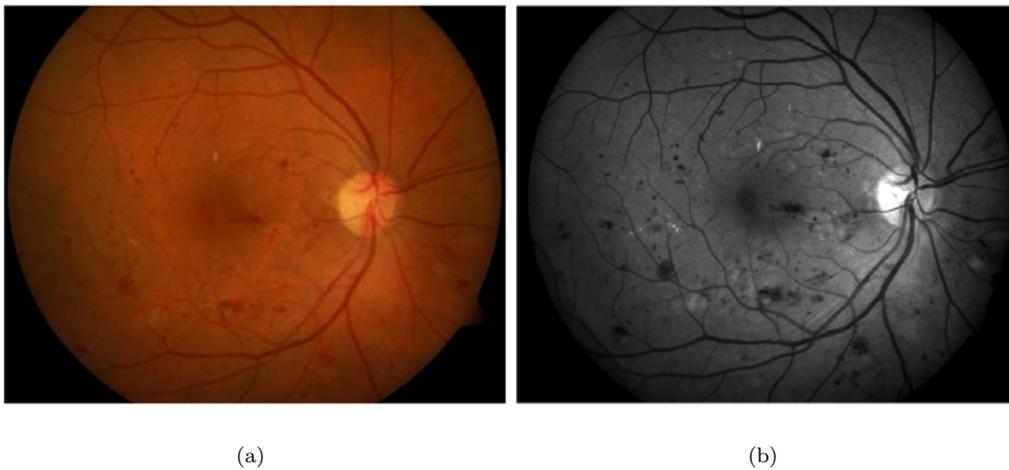


Figure 2.10: Examples of eye fundus images: (a) colour image of an eye fundus; (b) corresponding red-free image.

ALTERNATE DIAGNOSTIC MODALITIES

In addition to clinical eye examination and eye fundus photography, the fluorescein angiography and optical coherence tomography play an important role in the diagnosis of diabetic retinopathy. In the fluorescein angiography [104], a fluorescent dye (sodium fluorescein) is injected in the systemic circulation of a patient and by emitting light into patient's eye in specific wavelength the fluorescent properties of the dye are activated. The emitted light excites the dye molecules into the higher energy level and as the molecules return to the original state they emit lower energy light that is captured using eye fundus photography. The obtained image is called angiogram. Since the dye circulates in the ocular vasculature, the fluorescein angiography provides valuable information for the diseases pertaining retinal vasculature such as microaneurysms, capillary nonperfusion and vessels leaking exudate in diabetic retinopathy. Minor disadvantages are the injection and in rare occasions side effects such as nausea. It is worth noting that automatic image analysis algorithms can be applied to the fluorescein angiograms obtained using digital eye fundus photography. A fluorescein angiogram is shown in Fig.2.11.



Figure 2.11: Examples of eye fundus images: (a) colour image of an eye fundus; (b) corresponding fluorescein angiogram.

Optical coherence tomography (OCT) produces a two-dimensional cross-sectional image of ocular tissue structures, where the dimensions are propagation direction of the light and the perpendicular spatial direction [78]. A broadband beam of light (laser) is scanned across the ocular tissue and due to transparent structures of the retina the time of propagation is longer from the deeper tissue layers. Optical coherence tomography collects the reflected light and use the low-coherence interferometer to measure the time-of-flight delay [104]. The final optical coherence tomography image is composed from several axial scans and using several OCT images a computational three-dimensional reconstruction of

the retina can be devised. In diabetic retinopathy it is mainly used to provide accurate information about macular swelling and its type [123].

Other modalities used in eye examination [78, 172]: adaptive optics ophthalmoscopy, colour Doppler imaging, computed tomography, confocal laser scanning microscope, magnetic resonance, ophthalmic ultrasound, retinal thickness analyser and scanning laser polarimetry.

2.4 Screening diabetic retinopathy

The prevention of diabetic retinopathy concentrates on controlling the complications of diabetes in the eye through lifestyle and early treatment. These preventive actions can severely delay or stop the progression of the disease, prevent blindness and improve the quality of life. Since there are no salient symptoms in the early stages of diabetic retinopathy, and the number of symptoms and severity predominantly increase with time, a cost-effective screening over large populations is required [93, 145]

Screening is a secondary preventative action which aims to find and treat conditions that have already occurred, but which have not reached a stage that require medical attention. For successful screening the following criteria should be met [187]:

- (P1) “The condition sought should be an important health problem.”
- (P2) “There should be an accepted treatment for patients with recognized disease”
- (P3) “Facilities for diagnosis and treatment should be available.”
- (P4) “There should be a recognizable latent or early symptomatic stage.”
- (P5) “There should be a suitable test or examination.”
- (P6) “The test should be acceptable to the population ”
- (P7) “The natural history of the condition, including development from latent to declared disease, should be adequately understood.”
- (P8) “There should be an agreed policy on whom to treat as patients.”
- (P9) “The cost of case finding (including the diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as whole.”
- (P10) “Case-finding should be a continuing process and not a “once and for all” project.”

To shortly review the current state of screening diabetic retinopathy, the Finnish diabetic population is used as a case study for the screening criteria. In Finland, which has a population of around 5 million, there are 280 000 people under diabetes care and this is expected to double in every 12 years if major preventive actions are not undertaken [189]. In addition, it has been estimated that 200 000 people are undiagnosed [189]. Since each of the diabetic patients may lose their sight due to diabetic retinopathy, it can be considered an important health problem (P1). The early signs (P4) and progressive nature (P7) of diabetic retinopathy is well documented as well as the severity scales (P8) (Section 2.2.1), and for the diagnosis there are available (P3), acceptable (P6) and non-invasive diagnostic modalities (Section 2.3). If diabetic retinopathy is diagnosed, it can be treated with laser treatment, medical therapy or surgical intervention (P2, P3) [107]. However, diabetes contributes to 15% of the total health care costs of which

90% is spent on treating complications such as diabetic retinopathy [28]. Clearly, the cost is not in balance with the expenditure on Finnish medical care as a whole due to the late treatment of diabetes (P9) and as a result the screening cannot be continuous (P10) if the number of diabetic patients doubles as it is estimated.

The problem lies in the process of grading the eye fundus images which is time consuming and repetitive, and requires attention of an ophthalmologist. The grading time is mostly spent on eye fundus images that are not leading to any medical action that is tedious and makes the diagnosis prone to errors. Moreover, the disease progress is reviewed at least once in 1-3 years which results an increasing amount of information for the examination. Thus, a large amount of the costs is tied to the people conducting the diagnosis.

Digital imaging of the eye fundus and automatic or semi-automatic image analysis algorithms provide a potential solution. By automating the grading process more patients could be screened and referred for further examinations, and the ophthalmologists would have more time for patients which require their attention.

2.5 Automatic detection of diabetic retinopathy

The interest towards automatic detection of diabetic retinopathy has been increasing along with the rapid development of digital imaging and computing power. However, the single most important event that attracted the wider attention of medical research community has been the decision to recognise digital imaging as an accepted modality to document eye fundus. Since then, a considerable amount of effort has been spent on automated detection and diagnosis of diabetic retinopathy from digital eye fundus images. The relevant research is well documented in three recent surveys [188, 127, 38] which encapsulate the main algorithms used in the field during the past 10-15 years. In this section, the methodology behind these existing approaches are shortly reviewed.

The review is divided into three parts of which the first two parts discuss the automatic detection of diabetic retinopathy from the lesion point of view (i.e. detecting lesions indicative of diabetic retinopathy) while the discussion in the third part concentrates on algorithms that attempt to detect the presence or even the severity of diabetic retinopathy. The review provides a short description for each method and the reported performances are gathered into summary tables. It should be noted, however, that the interpretation of the used performance measures varied between publications and different data sets were employed in the evaluation. Therefore, the direct quantitative comparison should not be made. Moreover, the performance measures are restricted to sensitivity, specificity, positive predictive value and average number of false positives which are discussed in Section 3.5.

MICROANEURYSMS AND HAEMORRHAGES

Some of the first automated detection methods for diabetic retinopathy were published by Baudoin et al. [18] to detect microaneurysms from fluorescein angiograms. By using a morphological top-hat transform with linear structuring element at different orientations small round shaped microaneurysms were distinguished from connected elongated

structures such as vessels. Although the top-hat transform was very sensitive to microaneurysms, it introduced too many false alarms. Spencer et al. [156] exploited this feature and used the top-hat transform to produce candidate microaneurysms. The true microaneurysms were then pruned by using post-processing based on their earlier work [157] and classification. The candidate microaneurysm segmentation was conducted using a combination of top-hat transform and matched filtering with region growing. To improve the sensitivity of the candidate search a shade correction and dynamic range normalisation steps were introduced in the pre-processing. After detection and segmentation of the candidate microaneurysms, the true microaneurysms were pruned from the spurious responses using a rule-based classifier with a number of shape and intensity based features. By using the computer vision based detection concept (i.e. image acquisition, pre-processing, candidate object segmentation and classification) Spencer et al. achieved a better control over the problem and allowed the easier development of variant methods [25, 50, 106, 57]. The main difference between the method proposed by Spencer et al. and the variant methods lay in the classification step, where different classifiers and features were used. The feature and classification selection was also studied by Ege et al. [32].

The intravenous use of the fluorescein restricts the use of fluorescein angiography in large scale screening that turned the interest of researchers towards the red-free and colour eye fundus photography. Unlike in the fluorescein angiograms, the microaneurysms appear dark in the red-free and colour eye fundus images, and have lower contrast. Otherwise the detection task is, however, quite similar. Based on the research in [25, 156], a version of the top-hat transform based method was presented for red-free images by Hipwell et al. [74], and for colour eye fundus images by Yang et al. [195] and Fleming et al. [45]. The top-hat approach was also studied in detection of haemorrhages by Fleming et al. [44].

An alternative mathematical morphology based approach was proposed by Klein et al. [178, 179, 181, 31] to overcome a shortcoming of the top-hat based methods: the linear structuring element at discreet orientations tended to detect tortuous vessels as candidate microaneurysms. Instead of using linear structuring element, a bounding box closing was applied with the top-hat transform. Since the candidate object detection of the top-hat transform based methods also missed number of true microaneurysms, Niemeijer et al. [117] proposed a red lesion (microaneurysm and haemorrhage) detection algorithm by introducing a hybrid method to relax the strict candidate object size limitations. A combination of the top-hat based method described in [25, 156] and a pixel-based classification scheme was proposed to produce a more comprehensive set of candidates. After detecting candidates the true red lesions were pruned in k-nearest-neighbour classification.

There are also number of approaches for microaneurysm detection published in literature that are not based on morphological operations. One of the first approaches applied in detection diabetic retinopathy was proposed by Gardner et al. [55] who conducted preliminary experiments to study whether neural networks can be used in screening diabetic retinopathy. The neural network and supervised learning was utilised on red-free eye fundus images to extract the microaneurysm and haemorrhage characteristics from set of image patches. Using the trained neural network the microaneurysms and haemorrhages were located from previously unseen set of test images. Kamel et al. [83] proposed a similar method for fluorescein angiograms to substitute the slow intermediate

filtering sequences in the previously described top-hat transform based methods.

The approaches in [151, 173, 33, 73, 72, 102, 53] assumed that dark areas in the colour eye fundus images consist of vessels, microaneurysms and haemorrhages. By excluding the vessels and vessel segments, the remaining objects could be identified as microaneurysms and haemorrhages. Sinthanayothin et al. [151] and Usher et al. [173] used recursive region growing to cluster the dark areas in the image and classified the vessel and vessel segments from the region growing result using a neural network. Instead of using region growing, Grisan and Ruggeri [33] detected dark objects by clustering similar pixels with high local spatial density. The method was further improved by García et al. [53] who added an automatic feature selection and classification step (neural networks) to prune the true red lesions.

Pallawala et al. [125] proposed a different approach, where the microaneurysm detection was defined as a segmentation problem between micro-regions (i.e. microaneurysms) and the background regions (i.e. other eye fundus structures). The graph theory-based segmentation procedure exploited the similarity and spatial proximity of image pixels to cluster small tightly grouped areas into one category and more loosely connected larger areas into the other. In the post-processing step, the true microaneurysms were pruned from the small tightly grouped areas using the intensity difference between the area and the immediate surrounding.

Xiaohui and Chutatape [198, 199] extracted the characteristic features of haemorrhages from image templates using the principal component analysis (PCA). The extracted features were used with the support vector machine to classify the image patches of previously unseen colour eye fundus image. To detect the different sized haemorrhages, a pyramid of images were generated by computationally changing the image resolution. Quelled et al. [130, 131] also used image templates in detection of microaneurysms by conducting template matching in wavelet domain. Bhalerao et al. [20] modelled the circular shape and locally darker appearance of microaneurysms by combining the Laplacian of Gaussian and circular-symmetry operator in filtering the candidate microaneurysms. A template from each candidate microaneurysm location was extracted to find the true microaneurysms using a PCA-based region analysis.

An alternative approach was proposed by Narasimha-Iyer et al. [109] by introducing the use of temporal changes between colour eye fundus images to detect the symptoms of diabetic retinopathy. First, the images were pre-processed by registering the images with the dual-bootstrap iterative closest point algorithm, correcting the illumination using the iterative homomorphic surface fitting, and removing the dust particles using the ratio of the green and red reflectance components. The difference of two pre-processed eye fundus images were then classified to change and no-change regions, and further to lesion categories including microaneurysms and haemorrhages based on the difference ratio values.

Reported performances of the previously described microaneurysm and haemorrhage detection algorithms are shown in Tables 2.1 and 2.2. The algorithms that were not specifically designed for microaneurysm or haemorrhage detection, but which detected dark and reddish diabetic lesions, are denoted as red lesion detection algorithms in Table 2.3.

Table 2.1: Reported performances of microaneurysm detection algorithms in chronological order (within each sub-group), where SN = sensitivity, SP = specificity and PPV = positive predictive value and AFP = average number of false positives (FP).

	SN	SP	PPV	AFP	Train and test sets
Task: detect image pixels with microaneurysms					
Bhalerao et al. [20]	0.83	0.80	-	-	Tested against public DIARETDB1 image databases comprising total of 89 images.
Task: detect image segments with microaneurysms					
Ege et al. [32]	0.69	-	-	-	Tested on 161 image segments with microaneurysms extracted from 134 images.
Pallawala et al. [125]	0.93	-	-	-	Tested on 70 image segments with microaneurysms.
Task: detect connected image areas that represent microaneurysms					
Spencer et al. [156]	0.82	-	-	-	Tested on 4 images with 71 microaneurysms.
Cree et al. [25]	0.82	0.84	-	5.70	Trained on 68 images with 394 microaneurysms, tested on 20 images with 297 microaneurysms.
Frame et al. [50]	0.84	0.85	-	-	Trained on 68 images with 1659 objects (400 microaneurysms), tested on 20 images with 1067 objects (297 microaneurysms).
Hipwell et al. [74]	0.43	-	-	0.11	Tested on 88 images (62 with microaneurysms).
Yang et al. [195]	0.80	0.90	-	-	Tested on 3 images.
Walter and Klein [178, 179]	0.86	-	0.75	7.70	Tested on 5 images with 133 microaneurysms.
Quellec et. al. [130]	0.88	0.96	-	-	Trained and tested on database of 995 images
Quellec et. al. [131]	0.90	-	0.90	-	Tested on 35 green filtered colour photographs.
Quellec et. al. [131]	0.94	-	0.92	-	Tested on 50 fluorescein angiograms.
Quellec et. al. [131]	0.90	-	0.89	-	Tested on 35 colour photographs.
Walter et al. [181, 31]	0.89	-	-	2.13	Trained on 21 images (all images had at least 1 microaneurysm), tested on 94 images (68 with 373 microaneurysms).
Task: detect images with microaneurysms					
Hipwell et al. [74]	0.81	0.93	-	-	Tested on 3609 images with 913 microaneurysms.
Fleming et al. [45]	0.85	0.83	-	-	Tested on 1440 images with 356 microaneurysms

Table 2.2: Reported performances of haemorrhage detection algorithms in chronological order (within each sub-group).

	SN	SP	PPV	AFP	Train and test sets
Task: detect image segments with haemorrhages					
Gardner et al. [55]	0.74	0.74	-	-	Trained on 11694 image segments (5744 with haemorrhages), tested on 684 image segments (342 with haemorrhages), comprising 179 images in total
Ege et al. [32]	0.83	-	-	-	Tested on 87 image segments with haemorrhages extracted from 134 images.
Zhang and Chutatape [198, 199]	0.90	-	-	2	Trained on 1062 image segments (262 with haemorrhages) extracted from 30 images, tested on 15 images with 237 haemorrhages
Task: detect images with haemorrhages					
Hatanaka et al. [73]	0.85	0.21	-	-	Tested on 113 images.
Hatanaka et al. [72]	0.80	0.80	-	-	Trained on 20 images, tested on 125 images (35 with haemorrhages).

Table 2.3: Reported performances of red lesion (i.e. haemorrhages and microaneurysms) detection algorithms in chronological order (within each sub-group).

	SN	SP	PPV	AFP	Train and test sets
Task: detect image segments with red lesions					
Sinthanayothin et al. [151]	0.78	0.89	-	-	Tested on 30585 image segments (1015 with red lesions), comprising 30 images in total (14 with red lesions).
Task: detect connected image areas that represent red lesions					
Niemeijer et al. [117]	0.30	-	-	0.80	Trained on 50 images (28 with 510 red lesions), tested on 50 images (27 with 348 red lesions).
Grisan and Ruggeri [33]	0.94	-	-	-	Tested on 6 images with 108 red lesions.
García et al. [53]	0.86	-	0.71	-	Trained on 50 images with 4720 objects (2360 red lesions), tested on 50 images (25 with red lesions).
Task: detect images with red lesions					
Niemeijer et al. [117]	1.00	0.87	-	-	Trained on 50 images (28 with 510 red lesions), tested on 50 images (27 with 348 red lesions).
García et al. [53]	1.00	0.60	-	-	Trained on 50 images with 4720 image regions (2360 with red lesions), tested on 50 images (25 with red lesions).

HARD AND SOFT EXUDATES

The early work in the automated detection of hard and soft exudates generally investigated possibilities of thresholding techniques. Ward et al. [183] proposed a semi-automatic exudate detection system based on shade correction and thresholding, where the user interaction was required in the thresholding. By introducing a dynamic thresholding procedure Philips et al. [129] and Zheng et al. [200] provided a considerable improvement to the previous system. Philips et al. [129] detected large high intensity areas from red-free eye fundus images using a global thresholding scheme, whereas a block-wise local thresholding was applied to segment the smaller exudates. The method was able to produce relatively good results in detecting the exudate pixels, but at the same time an unacceptable number of false positives were generated. As a result, the ophthalmologic expertise was required in the interpretation of the detection results. To counter the false positives, Zheng et al. [200] introduced the use of local neighbourhood in the dynamic block-wise local thresholding procedure that was later adopted by Sagar et al. [140].

Simultaneously with the early thresholding, Goldbaum et al. [62, 61] proposed a bright lesion detection approach based on template matching and edge detection. By computationally resampling the image resolution and applying the template matching, the bright lesions of all sizes were located. The outlines of each located lesion were refined using edge detection. Most importantly, Goldbaum et al. [63] took into account the sub-classes of bright lesions, i.e. hard and soft exudates, that should be differentiated for the diagnosis of diabetic retinopathy. For the identification, a spherical colour space was introduced.

From template matching and thresholding, the method development turned towards supervised statistical pixel-based lesion classification. The first pixel- and block-based classification approaches utilised a simple minimum distance discriminant classifier to classify image pixels into two classes (i.e. to bright lesion and background) according to their pixel value. Wang et al. [182] applied the minimum distance discriminant classifier in the colour space proposed by Goldbaum et al. [63] to detect candidate hard exudate pixels. The true hard exudate pixels were then pruned using the contrast information of the local neighbourhood. Sánchez et al. [141] further developed the method using

alternative approaches for the non-uniform illumination correction and hard exudate pixel pruning. Goh et al. [60] applied the minimum distance discriminant classifier directly in RGB colour space to detect hard exudate pixels.

Niemeijer et al. [116] proposed a pixel classification scheme based on k-nearest neighbour classification to detect and differentiate hard and soft exudates, and drusen. The system searched for candidate bright lesion pixels according to the features selected in the training stage. By using the density of classified lesion pixels among the neighbouring pixels in the feature space, a lesion probability was assigned for each pixel in the test image. The high probability regions were pruned to find the true bright lesions by extracting descriptive features for each region and applying the KNN classification once again. Finally, a linear classifier was used to classify the detected true bright lesions to hard and soft exudates and drusen. To improve the hard exudate detection, Sanchez et al. [144] introduced the use contextual information, i.e. the spatial relations of the surrounding anatomical structures and similar lesions.

A feature-based classification scheme was also proposed by Xu and Luo [193], where a feature combination based on stationary wavelet transform and gray level co-occurrence matrix was used to characterise the textural properties of hard exudates. The pixel-level classification was conducted using a support vector machine. Another support vector machine-based approach was suggested by Silberman et al. [149], where the local textural properties were extracted using a scale-invariant feature transform (SIFT).

Given the irregular properties of hard exudates in shape and scale, the pixel- and block based clustering methods provided another strong approach for local segmentation of hard exudates. Hsu et al. [77] utilised dynamic clustering after shade correction to group the image pixels into (bright) lesion and non-lesion areas. The hard exudate clusters were differentiated from other bright lesions based on the contrast difference between the (bright) lesion and non-lesion classes. A more elaborate clustering technique was proposed by Osareh et al. [120, 119, 122], where the candidate object detection was based on Gaussian-smoothed histogram analysis and fuzzy c-means (FCM) clustering. The idea was to use the Gaussian-smoothed histogram analysis to coarsely cluster the image according to the significant extreme points in the image histogram. The FCM clustering was then used to assign the remaining unclassified pixels to the previously determined clusters. To identify the hard exudate regions from the non-hard exudate regions image features were extracted from the clustered regions and then classified using a neural network approach. Zhang and Chutatape [197, 199] further improved the FCM method to exploit information in the neighboring pixels to decrease the effect of the image noise.

Ram and Sivaswamy [132] studied a clustering scheme, where two feature spaces were composed by combining colour channels of different colour spaces. The objective was to construct two divergent feature spaces descriptive for different eye fundus structures. First, the k-means clustering was used on both feature spaces to group similar image pixels and then by using reasoning and the complementary behaviour of the feature spaces, a set of candidate hard exudate regions were selected from the clustering results. To suppress false positives, the colour information of the candidate regions before and after the colour channel decorrelation procedure was exploited.

Region growing is another clustering technique applied in detection of hard and soft

exudates. Luo et al. [102] extracted candidate bright lesions using a block-wise region-growing scheme after emphasising bright areas in each image block by manipulating histogram properties of L and u colour components in LUV colour space. The difference between the mean pixel value inside and along the region contour was applied in the search of true bright lesions. The method was later adopted by Li and Chutatape [99, 100]. The widely cited research of Sinthanayothin et al. [151] proposed another recursive region-growing procedure to identify similar pixels in the detection of bright lesions. After the pixels were merged into regions, the bright lesions were differentiated by thresholding. Prior to the region-growing procedure a local colour contrast enhancement step was performed. A more recent region-growing approach was presented by Esrawan et al [35], where the bright lesions were detected from edge magnitude image using a marker driven watershed transform. To emphasise the regions of interest, an average filtering and contrast stretching preprocessing steps were performed.

In addition to the classification and clustering methods, a number of mathematical morphology based approaches have been published in the literature. Unlike the previously described methods, Walter et al. [180, 179] attempted to morphologically reconstruct the eye fundus image exclusive of bright lesions. The method was based on the idea that the difference between the reconstructed image and the original image would ultimately express the bright lesion locations. Firstly, the bright candidate lesions were coarsely located based on their local contrast variation after morphologically suppressing the blood vessels. Secondly, the candidate bright lesion areas were removed and morphologically reconstructed to correspond the appearance of the retinal background. By thresholding the difference of the original image and the reconstructed image, the final lesion areas were obtained. The same idea was later applied and refined by Noronha et al. [118] and Welfer et al. [146].

Ravishankar et al. [133] applied morphological operations to suppress the blood vessel network and to emphasise the bright lesion boundaries by dilating the image at two different scales and subtracting the outcomes. By using dynamic thresholding and morphologic filling, the candidate exudates were extracted from the dilation result by finding the connected regions within the emphasised lesion boundaries. In the final stage, brightness and edge properties of the candidate regions were used with the rule-based classification to determine the true hard exudate regions. Kumari et al. [94] had a very similar approach without the final classification step and with the dynamic thresholding step substituted to edge detection with the Laplacian of Gaussians.

Fleming et al. [46] utilised their earlier experiences with the microaneurysm detection in [45] and introduced an inversion of the morphological top-hat technique to detect hard exudates. A linear structuring element at different orientations and scales were applied with the top-hat transform to emphasise bright regions in the image. By using a dynamic thresholding scheme the candidate exudates were extracted from the top-hat result and finally classified to exudates, drusen and background according to their colour, shape, brightness and the contrast difference to the local background. The use of contextual information in the form of distance to the closest microaneurysm was also studied.

An alternative hard exudate detection approach was proposed by Sánchez et al. [143, 142], where the statistical properties of an eye fundus image were captured into a mixture of Gaussians. By estimating the probability density function of each image and by analysing the concavities of the approximated density functions the candidate bright regions were

thresholded from the background. False positives were eliminated in the postprocessing step, where the edge strength of the candidate regions were made use of. A similar idea was introduced by García et al. [52, 54], where the mixture model was substituted to a global and several local image histograms, and the true hard exudate search to automatic feature selection and neural network classification.

Reported performances of the previously described bright lesion detection algorithms are shown in Table 2.4. The bright lesion detection algorithms are not specifically designed for hard or soft exudate detection, but which detected bright and yellowish diabetic lesions. Reported performances of the previously described hard and soft exudate detection algorithms are shown in Tables 2.5 and 2.6.

Table 2.4: Reported performances of bright lesion (i.e. soft and hard exudates) detection algorithms in chronological order (within each sub-group).

	SN	SP	PPV	Train and test sets
Task: detect pixels with bright lesions				
Eswaran et al. [35]	0.94	-	0.92	Tested against public STARE and DRIVE image databases.
Task: detect connected image areas that represent bright lesions				
Goldbaum et al. [62]	0.95	-	-	Tested against public STARE image database.
Niemeijer et al. [116]	0.95	0.88	-	Trained on 130 images with 3188 bright lesions, tested on 300 images (105 with bright lesions).
Zhang and Chutatape [197, 199]	0.97	0.96	-	Trained on 1494 objects with 511 bright lesions, tested on 692 objects with 260 bright lesions.

Table 2.5: Reported performances of soft exudate detection algorithms in chronological order (within each sub-group).

	SN	SP	PPV	Train and test sets
Task: detect image segments with soft exudates				
Ege et al. [32]	0.80	-	-	Tested on 20 image segments with soft exudates extracted from 134 images.
Task: detect connected image areas that represent soft exudates				
Niemeijer et al. [116]	0.70	0.93	-	Trained on 130 images (45 with soft exudates), tested on 300 images (30 with soft exudates).
Zhang and Chutatape [197, 199]	0.88	0.84	-	Trained on 1494 objects with 54 soft exudates, tested on 692 regions (47 with soft exudates)

DIABETIC RETINOPATHY DETECTION

The microaneurysm count was one of the first criterias proposed by Baudoin et al. [19] to differentiate healthy and diabetic images. The idea exploited the knowledge that the number of microaneurysms correlates with the severity of diabetic retinopathy. Since the idea was simple, it was considered as a promising candidate to detect the presence of diabetic retinopathy in eye fundus images which boosted the development of microaneurysm detection algorithms. In the screening of diabetic retinopathy, the use of microaneurysm count was later studied by Kahai et al. [80, 81].

Since other lesions indicative of more severe disease state can occur prior to microaneurysms, the automatic detection of diabetic retinopathy using microaneurysm count had room for improvement. To overcome this problem, different lesion counts were experimented e.g., Sinthanayothin et al. [152] devised the presence of diabetic retinopathy in

Table 2.6: Reported performances of hard exudate detection algorithms in chronological order (within each sub-group).

	SN	SP	PPV	AFP	Train and test sets
Task: detect pixels with hard exudates					
Phillips et al. [129]	0.87	-	-	-	Tested on 13 images with 30 hard exudate regions.
Walter et al. [179]	0.93	-	0.95	-	Not reported
Walter et al. [180]	0.93	-	0.92	-	Tested on 30 images (15 with hard exudates).
Osareh [119]	0.90	-	0.89	-	Tested on 40 images.
Osareh et al. [122]	0.94	-	0.92	-	Trained on 150 images (75 with hard exudates), tested on 150 images (75 with hard exudates).
Ram and Sivaswamy [132]	0.72	-	0.87	0.22	Tested against public DIARETDB1 image database comprising total of 89 images (38 with hard exudates).
Ravishankar et al. [133]	0.95	0.91	-	-	Tested against public STARE, DRIVE, DIARETDB0 and Red Atlas image databases comprising total of 516 images (345 with hard exudates).
Task: detect image segments with hard exudates					
Gardner et al. [55]	0.93	0.93	-	-	Trained on 11694 image segments (5744 with haemorrhages), tested on 684 image segments (342 with haemorrhages), comprising 179 images in total
Ege et al. [32]	0.99	-	-	-	Tested on 175 image segments with hard exudates extracted from 134 images.
Osareh et al. [120]	0.92	0.82	-	-	Trained on 3000 segments (1205 hard exudates), tested on 1037 segments (417 hard exudates), comprising 42 images in total.
Sinthanayothin et al. [151]	0.89	1.00	-	-	Tested on 60780 image segments (893 with hard exudates), comprising 30 images in total (21 with hard exudates).
Xu and Luo [193]	0.88	0.80	-	-	Trained and tested on 14 images (25 hard exudate regions).
Welfer et al. [146]	0.71	0.99	0.22	-	Tested against public DIARETDB1 image database using total of 47 images with hard exudates.
Task: detect connected image areas that represent hard exudates					
Sagar et al. [140]	0.99	-	0.94	-	Tested on 25 images (15 with 289 hard exudates).
Sánchez et al. [141]	0.80	-	-	3.2	Tested on 20 images.
Zhang and Chutatape [197, 199]	0.88	0.84	-	-	Trained on 1494 objects with 457 hard exudates, tested on 692 objects with 213 hard exudates
Sánchez et al. [143]	0.90	-	0.83	-	Tested on 20 images (10 with hard exudates).
Niemeijer et al. [116]	0.95	0.86	-	-	Trained on 130 images with 1113 hard exudates, tested on 300 images (42 with hard exudates).
García et al. [52]	0.84	-	0.63	-	Trained on 2380 image regions (1190 with hard exudates), tested on 50 images (25 with hard exudates).
Sánchez et al. [142]	0.90	-	0.97	0.63	Trained on 26 images, tested on 80 images (40 with hard exudates).
Task: detect images with hard exudates					
Wang et al. [182]	1.00	0.70	-	-	Tested on 154 images (54 with hard exudates).
Hsu et al. [77]	1.00	0.74	-	-	Tested on 543 images (31 with hard exudates).
Li and Chutatape [99, 100]	1.00	0.71	-	-	Tested on 35 images (28 with exudates).
Osareh [119]	0.96	-	0.89	-	Tested on 67 images (40 with hard exudates).
Sánchez et al. [143]	1.00	-	0.90	-	Tested on 20 images (10 with hard exudates).
Hatanaka et al. [73]	0.77	0.83	-	-	Tested on 109 images.
García et al. [52]	1.00	0.84	-	-	Trained on 2380 image regions (1190 with hard exudates), tested on 50 images (25 with hard exudates).
Fleming et al. [46]	0.95	0.85	-	-	Trained on 379 images (139 with hard exudates), tested on 13219 images (300 with hard exudates).
Sánchez et al. [142]	1.00	0.90	-	-	Trained on 26 images, tested on 80 images (40 with hard exudates).
Ravishankar et al. [133]	0.96	0.94	-	-	Tested against public STARE, DRIVE, DIARETDB0 and Red Atlas image databases comprising total of 516 images (345 with hard exudates).
Osareh et al. [122]	0.96	0.95	-	-	Trained on 150 images (75 with hard exudates), tested on 150 images (75 with hard exudates).
Silberman et al. [149]	0.87	0.87	-	-	Trained on 3390 image patches (1390 with hard exudates), tested on 1000 images (87 with hard exudates).

eye fundus images using the hard exudate count, Larsen et al. [96] and Hansen et al. [71] used microaneurysm and haemorrhage counts, and Usher et al. [173] used microaneurysm, haemorrhage and hard exudate counts. Reza and Eswaran [136] also attempted to devise the severity of diabetic retinopathy using a lesion count-based approach, whereas Estabridis and Figueiredo [34] used the spatial lesion locations.

Alternatively, machine learning is utilised in the literature. Goldbaum et al. [62] proposed a method, where an image-description language was utilised to symbolise the properties of lesions and normal structures in eye fundus images. By using a neural network and a set of representative images of known diagnosis, the connection between the image-description information and the known diagnosis was learned. Based on the training data, the trained network was able to detect images with diabetic retinopathy using the image-description information derived from the result of automatic eye fundus image analysis system. Instead of using image descriptions, Yagmur et al. [194] used neural networks with wavelet coefficients, Anitha et al. [15] trained the network using cluster centroids obtained from channel-wise fuzzy c-means clustering, and the work in [26, 111, 196] utilised shape related features extracted from the segmented vessel and lesion regions.

Another feature-based classification scheme based on machine learning was proposed by Agurto et al. [12, 17], where the normal and pathological structures were characterised according to their amplitude and frequency content, conceptually, describing the textural properties. After decomposing the image using the amplitude modulation - frequency modulation (AF-FM) decomposition and extracting the textural features for each image from several image locations, k-means clustering was used to cluster the extracted features into several groups. The number of textural features in each of the group defined the final descriptor for each image. Using the described descriptors and a partial least square classifier, the presences of diabetic retinopathy was detected in eye fundus images.

Goh et al. [59] combined several classifiers to detect diabetic retinopathy from eye fundus images by following the main principal of classifier ensembles, i.e. while the single best performance is achieved using the best classifier, the misclassified samples by the best classifier are not necessarily misclassified by rest of the classifiers. A set of nine ensembles containing several classifiers were constructed, where three classifier ensembles were used for the background, four for the blood vessel and three for the bright lesion detection. The classifiers within each ensemble were based on neural networks. By using decision rules, the classifier outputs within each ensemble and the resulting ensemble outputs were combined into single label representing background, blood vessel or bright lesion. Each sub-image in the image was assigned with such label and based on the image patches containing bright lesions the presence of diabetic retinopathy was finally detected.

Niemeijer et al. [112] also studied the use of multiple information sources and their fusion to achieve more robust detection for screening diabetic retinopathy. The proposed system analysed two images per eye and four images in total to produce a subject-wise score value that indicated the likelihood that the test subject needs to be examined by a human observer. Each of the four images were separately analysed with complex computer aided detection or diagnosis (CAD) system to determine the image quality and the diabetic lesion locations (red and bright lesions). The subject-wise score value was then combined from the outputs of the CAD systems and based on the score value the diabetic retinopathy was detected in eye fundus images. Fleming et al. [43] had a similar idea,

where eye fundus images of both eyes were first analysed for microaneurysms, haemorrhages, hard exudates and image quality and by using the output several image-based lesion measures were computed and combined into overall subject-wise score. Similarly as previously, the score value was used to detect the patients with referable diabetic retinopathy.

Reported performances of the previously described diabetic retinopathy detection algorithms are shown in Table 2.7.

Table 2.7: Reported performances of diabetic retinopathy detection algorithms in chronological order (within each sub-group).

	SN	SP	Train and test sets
Task: detect images with diabetic retinopathy			
Sinthanayothin et al. [152]	0.80	0.70	Tested on 771 images. (238 with DR)
Usher et al. [173]	0.87	0.67	Tested on 1406 images (427 with DR)
Kahai et al. [81]	1.00	0.67	Tested on 143 images.
Agurto et al. [12, 17]	0.92	0.52	Trained on 188 images (118 with DR), tested on 188 images (118 with DR). Public MESSIDOR image database were used.
Barriga et al. [17]	0.98	0.67	Tested against public MESSIDOR image database.
Goh et al. [59]	0.91	0.92	Tested on 1000 images (750 with DR). Public DIARETDB0 and DIARETDB1 were used in the evaluation
Task: detect subjects with diabetic retinopathy			
Larsen et al. [96]	0.90	0.80	Tested on 129 subjects (81 with DR).
Usher et al. [173]	0.95	0.53	Tested on 773 subjects (267 with DR)
Hansen et al. [71]	0.90	0.86	Tested on 83 subjects without pupil dilation (69 with DR)
Hansen et al. [71]	0.97	0.75	Tested on 83 subjects with pupil dilation (69 with DR)
Niemeijer et al. [112]	0.93	0.60	Tested on 15 000 subjects (394 with DR)
Fleming et al. [43]	0.97	0.50	Tested on 7586 subjects (1253 with DR)

2.6 Summary

In this chapter, diabetes-related eye diseases and their symptoms were discussed. For the most common diabetic eye disease, diabetic retinopathy, the diagnostic procedures and modalities were described. Since the eye fundus photography is non-invasive, efficient, easy to use and enable the documentation of the diagnostic data, it was considered the preferred tool for diagnosing diabetic retinopathy. The diagnosis of diabetic retinopathy from an eye fundus image was based on identifying the described symptoms of non-proliferative, proliferative retinopathy and maculopathy,

The chapter also addressed that future prospects of diabetic retinopathy are alerting and preventive actions, such as a large scale screening of high-risk individuals, are required. The number of diabetic patients at the current increase rate will eventually exceed the current screening capabilities due to the time consuming and tedious grading process of the eye fundus images. Fortunately, the development of digital imaging in the diagnosis of diabetic retinopathy, i.e., digital eye fundus photography, has provided a possibility to automate the grading process and thereby improve the efficiency of the screening. The fields of image processing and computer vision has shown a great interest towards automatic and semi-automatic detection of diabetic retinopathy, and number of publications are already presented in the literature as stated in the literature review. The need for automatic image analysis system to detect the early signs of diabetic retinopathy (haemorrhages, microaneurysms, hard and soft exudates) is the main motivation for this thesis.

Benchmarking of eye fundus image analysis algorithms

Public benchmarking image databases are an essential resource in the development of eye fundus image analysis algorithms that considerably help the medical imaging researchers to evaluate and compare state-of-the-art methods. Eventually, it leads to the development of better algorithms, and consequently, will support technology transfer from research laboratories to clinical practise. In this chapter, a benchmarking framework is described that provides guidelines on how to construct benchmarking image databases for eye fundus image analysis algorithms. The benchmarking framework comprises three mandatory requirements: 1) true patient images, 2) ground truth, and 3) an evaluation protocol. The guidelines in the benchmarking framework are utilised to establish the DIARETDB1 database to promote automated detection of diabetic retinopathy .

3.1 Introduction

Image databases and expert ground truths lie in the regular use of eye fundus image processing. However, the public availability of image databases is limited by the privacy protection of the patient information, and therefore reliable comparisons and state-of-the-art surveys are difficult to perform. In contrast to, for example, biometrics including face, iris, and fingerprint recognition, the research has been driven by public databases and solid evaluation protocols. Due to the continuous pressure to develop better algorithms, the databases have been extended and revised. The adoption of these research practises would be a valuable asset in the eye fundus image processing and analysis as well.

For every application it should be an acknowledged scientific contribution to provide a set of images, collect accurate and reliable ground truth, and devise a meaningful evaluation protocol. Once this pioneering work has been done, it sets an evaluation standard for the specified problem. The development of such benchmarking database from the scratch is demanding, laborious and time consuming, but fortunately certain tasks occur repeatedly and are reusable as such. In this chapter, the practical issues related to the construction of benchmarking databases are discussed and repeatedly occurring sub-tasks are pointed out. The given results and guidelines are applied to construct a revised version of the

diabetic retinopathy database DIARETDB1. The content of this chapter is reported in [85, 86].

Thacker et al. [163] have studied the performance characterisation of computer vision methods. Their study is comprehensive and general, and they provide good examples which are easily transferable to applications of medical image processing. The results from [163] can be utilised in every step of the method development, but in this chapter the analysis is restricted to the final diagnosis, i.e., the subject-wise decision making directly serving the clinical work. In other words, the framework omits the development and research phase evaluations, and construct the good practices to evaluate the performance of eye fundus image analysis algorithms. For that purpose, the eight general considerations adopted from [163] are addressed, and referred as the key questions:

- C1: “HOW IS TESTING CURRENTLY PERFORMED?” If a commonly used database and protocol are available, their validity for the development and evaluation needs to be examined. In the worst case, a new database needs to be constructed for which the proposed framework can be useful.
- C2: “IS THERE A DATA SET FOR WHICH THE CORRECT ANSWERS ARE KNOWN?” Such a data set can be used to report the results in accordance to other studies. This enables method comparison.
- C3: “ARE THERE DATA SETS IN COMMON USE?” See C1 and C2. Common data sets facilitate fair method comparison.
- C4: “ARE THERE EXPERIMENTS WHICH SHOW THAT ALGORITHMS ARE STABLE AND WORK AS EXPECTED?” These experiments can be realised if representative data and expert ground truth is available.
- C5: “ARE THERE ANY STRAWMAN ALGORITHMS?” If a strawman algorithm is included in the database it defines the baseline performance for other methods.
- C6: “WHAT CODE AND DATA ARE AVAILABLE?” By publishing the method’s code or at least executable version of it, other research groups can avoid repeating the same work.
- C7: “IS THERE A QUANTITATIVE METHODOLOGY FOR THE DESIGN OF ALGORITHMS?” This depends on the medical problem, but the methodology can be typically devised by following corresponding clinical work and practises. Understanding of the medical practitioners’ task which should be assisted or automated provides a conceptual guideline. If the database is correctly build to reflect the real world conditions, then the database implicitly reflects the applicability of the algorithm’s design to the problem.

C8: “WHAT SHOULD WE BE MEASURING TO QUANTIFY PERFORMANCE? WHICH METRICS ARE USED?” At least in the image-wise (subject-wise) experiments, the receiver operating characteristic (ROC) curve is in accordance with the medical practise, where the sensitivity and specificity values are in common use. The ROC curve, also known as ROC analysis, is a widely used tool in medical community for visualising and comparing methods based on their performance [201]. It is a graphical representation that describes the trade-off between the sensitivity and specificity (e.g., correctly classified normal images vs. correctly classified abnormal images). In the curve, the x -axis is defined as 1-specificity, and the y -axis is directly the sensitivity [39].

In general, $C1 \in C2 \in C3$, which means that if there is a commonly used data set in the form of, e.g., a benchmark database, the answers to $C1$ and $C2$ are known. Similarly, $C4 \in C5 \in C6$ which defines the maturity of the existing solutions. In the case where the data and code are both available and have been shown to work by achieving the required sensitivity and specificity rates, the solution is at a mature level and true clinical experiments can be started. $C7$ is a general guideline for the design to find an acceptable work flow for a specific problem and $C8$ sets the quantitative and meaningful performance measures.

The chapter is organised as follows. Section 3.2 presents the benchmarking requirements for eye fundus image analysis algorithms. Section 3.3 briefly surveys the public eye fundus image databases with respect to the key questions. Section 3.4 describes on how to collect patient images and ground truth. Section 3.5 discusses the use of ROC analysis in performance evaluation of eye fundus image analysis algorithms. Section 3.7 describes the benchmarking databases collected during the course of the work, where the given results and tools are used to establish the benchmarking database DIARETDB1. A baseline algorithm for the DIARETDB1 database is specified in Section 3.6, and the summary is given Section 3.8.

3.2 Requirements for benchmarking

Benchmarking image databases in eye fundus imaging require three mandatory components: 1) true patient images, 2) ground truth from experts, and 3) an evaluation protocol. Additional components, such as a baseline algorithm, provide notable additional value, but in the following, the three mandatory components are discussed.

3.2.1 True patient images

True patient images carry information which is meaningful for solving a given problem, i.e., algorithms which work with these images are expected to perform well also in practise. The images can be recorded using alternative subjects, such as animals that are genetically close to humans, and disease-related lesions can be produced artificially by using various substances. These are standard practises in medical research, but before drawing any general conclusions, their relevance and accuracy to the real world must be carefully verified. With true patient images, the results are biased by the distribution of database images with respect to the specific real population. The collection and selection

of images is further discussed in Sec. 3.4. The true patient image requirement concerns the key questions *C2*, *C3*, *C4*, and *C6*.

3.2.2 Ground truth given by experts

Ground truth must be accurate and reliable in the sense that it is statistically representative over experts. In the field of eye fundus image processing, it is advisable that the tools for ground truth annotation are provided by computer vision scientists, but the images are selected and annotated by medical experts specialised in the specific field. It is recommended to state the level and field of specialisation, e.g., senior and junior ophthalmologists. Information on the level can be useful later if the database is used for educational purposes. For example, junior doctors or students manually inspect the images and their annotations are compared to the expert-verified ground truth. It is also clear that the ground truth must be independently collected from multiple experts. This can be laborious and expensive, but it enables statistical studies of reliability. In the case of multiple experts, disambiguation of the data is often necessary prior to the application of machine learning methods. The requirement of collecting the ground truth from the experts concerns the key questions *C2*, *C3*, *C4*, and *C6*.

3.2.3 Evaluation protocol

A valid evaluation protocol providing quantitative and comparable information is essential for reliable performance evaluations. Most articles related to eye fundus image analysis report the sensitivity and specificity separately, but they are meaningless metrics until a method can produce superior values for both. Following the mainstream in the recent computer vision literature, the proposed evaluation protocol utilise the ROC analysis. The approach is essentially the same as reporting the sensitivity and specificity, but provides the evaluation result over all possible combinations of these values. It turns out that in benchmarking, the comparison of ROC curves is problematic, and therefore, specific well-justified operation points or area under the curves can be used as a single measure. This issue is further discussed in Sec. 3.5. In addition to the evaluation protocol, a baseline method (*C5*) or at least the results with the baseline method are helpful since they set the base level which any new methods should outperform. From another viewpoint, the best reported results by using a commonly accepted database set the standard norm. The evaluation protocol requirement concerns the key questions *C1*, *C4*, *C7*, and *C8*.

3.3 Eye disease databases

This section describes the most important public benchmarking databases in eye fundus image analysis. The database review provides a short description for each database, where the key questions *C1* – *C8* addressed in Section 3.1 are used to highlight the main properties. Since each database is publicly available they are expected to be in common use (*C3*).

STARE (Structured analysis of the retina) [9] is one of the most used reference image database in literature (*C3*, *C4*) for comparing blood vessel detection and optic disc

localisation algorithms. The STARE website [9] provides 20 images with pixel-wise hand-labeled ground truth for blood vessel detection (*C2*), and 81 images for optic disc localisation without ground truth. The performance of blood vessel detection is measured using the ROC curve analysis, where the sensitivity is the proportion of correctly classified blood vessel pixels and the specificity is the proportion of correctly classified normal pixels (*C8.1*) [76]. In the evaluation of optic disc localisation, the proportion of correctly localised optic discs indicates the performance and the localisation is successful if the centre of optic disc generated by the algorithm is within 60 pixels from the ground truth (*C8*) [75]. The evaluation procedures for both data sets are published with vessel detection algorithm and baseline results (*C5*) [75, 76].

DRIVE (Digital retinal images for vessel extraction) [5, 158] is another well-known reference database for blood vessel detection (*C3*), which contains 40 retinal images (*C6.2*) with manually segmented pixel-wise ground truth (*C2, C6.2*). The manual segmentation task was divided between three medical experts, and the database was published along with vessel detection algorithm (*C5*) [158]. The detection performance is measured similarly as in the STARE database, i.e., comparing the sensitivity to the specificity (*C8.1*) from which the area under curve (AUC) is computed to produce the final measure for the algorithm comparison (*C8.2*) [5, 158]. In addition, the authors implemented and internally evaluated a number of blood vessel detection algorithms from various research groups and the results were published in [114] and on the DRIVE database website (*C4*) [5].

MESSIDOR (Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology) [7] is a reference image database collected to facilitate computer-assisted image analysis of diabetic retinopathy. Its primary objectives are to enable evaluation and comparison of algorithms for analysing the severity of diabetic retinopathy, prediction of the risk of macular oedema, and indexing and managing image databases, i.e., support image retrieval. For the evaluation, the MESSIDOR database website [7] provides 1200 images (*C6.2*) with image-wise severity gradings (*C2, C6.2*) from three ophthalmologic departments including descriptions for the severity grading. It is noteworthy to mention that the severity gradings are based on the existence and number of diabetic lesions, and their distance from the macula.

CMIF (Collection of multispectral images of the fundus) [2, 160] is a public multispectral retinal image database. The spectral images were obtained by implementing a “filter wheel” into a fundus camera containing a set of narrow-band filters corresponding to the set of desired wavelengths [160]. The database itself consists of normal and abnormal images (*C6.2*) spanning a variety ethnic backgrounds covering 35 subjects in total [160]. As such, the database is not ready for benchmarking, but it provides a new insight into retinal pathologies.

ROC (Retinopathy online challenge) [8, 115] follows the idea of asynchronous online algorithm comparison proposed by Scharstein and Szeliski [147] for stereo correspondence algorithms (Middlebury Stereo Vision Page), where a web evaluation interface with public evaluation data sets ensures that the submitted results are comparable. The research groups download the data set, submit their results in the required format, and the results are evaluated by the web evaluation system. Since the evaluation is fully automatic,

the research groups can submit and update their results continuously. In the current state, the ROC database website [8] provides 100 retinal images (C6.2), a ground truth (C2, C6.2) and an online evaluation system for microaneurysms, and the evaluation results for a number of detection algorithms (C4). The algorithm performance is measured by comparing the sensitivity (the proportion of correctly classified lesions) against the average number of false positives in the image, i.e., free-response receiver operating characteristic curve (FROC) (C8.1) [115]. The sensitivities of predefined false positive points are averaged to generate the final measure for algorithm comparison (C8.2) [115]. The annotations were gathered from 4 medical experts by marking the location, approximate size, and confidence of the annotation. Consensus of two medical experts was required for a lesion to be selected the ground truth.

REVIEW (Retinal vessel image set for estimation of widths) [10, 13] is a new reference image database to assess the performance of blood vessel width measurement algorithms. To characterise the different vessel properties encountered in the retinal images, the database consists of four image sets: 1) high resolution image set (4 images); 2) vascular disease image set (8 images); 3) central light reflex image set (2 images), and 4) kick point image set (2 images) (C6.2). The REVIEW database concentrates on high precision annotations, and therefore, it provides only segments of blood vessels and not the whole network. To achieve high precision, the human observers used a semi-automatic tool to annotate a series of image locations from which the vessel widths were automatically determined [13]. The annotations were gathered from three medical experts, and the mean vessel width was defined as the ground truth (C2, C6.2). In the evaluation, the performance is measured using an unbiased standard deviation of the width difference between the algorithm-estimated vessel widths and the ground truth (C8) [13].

In general, most of the reference databases reach the minimal requirements for benchmarking image analysis algorithms, i.e., they provide true patient images, ground truth from experts, and an evaluation protocol (Table 3.1). In some cases, the usability is already at a mature level, e.g., in the case of the web evaluation system in the ROC database. The primary shortcomings appear to be related to the availability of software (C6.1) and how the algorithm’s design for the medical problem is observed (C7). By publishing the method’s code, other research groups could avoid repeating the same work and if the database is correctly build to reflect real world conditions, then the database implicitly reflects the applicability of the algorithm’s design to the problem. The database properties in terms of the key questions are summarised in Table 3.1 and for comparison the proposed DIARETDB1 database properties are summarised in Table 3.2.

3.4 Patient images and ground truth

3.4.1 Collecting patient images

The task of capturing and selecting patient images should be conducted by medical doctors or others specifically trained for the task. For a selected image set, there are two issues which should be justified: 1) distribution correspondence with the desired population and 2) privacy protection of patient data.

In DIARETDB1, the ophthalmologists wanted to investigate the accuracy of automatic methods analysing fundus images of patients who are in a serious risk of having diabetic

Table 3.1: Summary of the current state of the reference image databases in terms of the key questions addressed in Section 3.1.

Key questions	STARE (vessel)	STARE (disc)	DRIVE	MESSIDOR	CMIF	ROC	REVIEW
<i>C2</i> : “Is there a data set for which the correct answers are known?”	x		x	x		x	x
<i>C3</i> : “Are there data sets in common use?”	x	x	x	x	x	x	x
<i>C4</i> : “Are there experiments which show algorithms are stable and work as expected?”	x		x			x	
<i>C5</i> : “Are there any strawman algorithms?”	x	x	x				
<i>C6.1</i> : “What code is available?”						x	
<i>C6.2</i> : “What data is available?”	x	x	x	x	x	x	x
<i>C7</i> : “Is there a quantitative methodology for the design of algorithms?”							
<i>C8.1</i> : “What should we be measuring to quantify performance?”	x	x	x			x	x
<i>C8.2</i> : “What metrics are used?”		x	x			x	x
$\sum x$	6	5	7	3	2	7	5

retinopathy. Consequently, the images do not correspond to the actual severity or prevalence of diabetic retinopathy in the Finnish population, but to provide clear findings for automated detection methods. The data is, however, clinically relevant since the studied sub-population is routinely screened by Finnish primary healthcare. It is important to include this information to the database description to prevent misunderstandings and to justify the applicability of the results.

The privacy protection of patient data is a task related to the ethics of clinical practise, medical research, and also to data security. A permission for collecting and publishing the data must be acquired from a corresponding national organisation (e.g., national ethics committee of medical research) and from the patients themselves. Moreover, all data must be securely stored, i.e. all patient information, such as identifying metadata, must be explicitly removed from images which are to be used in a public database. In

Table 3.2: Summary of the DIARETDB1 V2.1 database in terms of the key questions addressed in Section 3.1.

Key questions	DIARETDB1 V2.1
<i>C2</i> : “Is there a data set for which the correct answers are known?”	Yes.
<i>C3</i> : “Are there data sets in common use?”	Yes (publicly available at [4]).
<i>C4</i> : “Are there experiments which show algorithms are stable and work as expected?”	Experimental results are reported for the baseline algorithm presented in Section 3.6.
<i>C5</i> : “Are there any strawman algorithms?”	No, but the baseline algorithm sets the baseline results for the DIARETDB1 database.
<i>C6.1</i> : “What code is available?”	Functionality for reading/writing images and ground truth, baseline algorithm, and annotation software (publicly available at [4, 6]).
<i>C6.2</i> : “What data is available?”	Images and ground truth (XML) (publicly available at [4]).
<i>C7</i> : “Is there a quantitative methodology for the design of algorithms?”	No, but medical practise is used as a guideline at each development step.
<i>C8.1</i> : “What should we be measuring to quantify performance?”	Image- and pixel-based ROC analysis (description in Section 3.5).
<i>C8.2</i> : “What metrics are used?”	Equal error rate (EER) defined in Section 3.5.

DIARETDB1, the eye fundus images were acquired using a standard fundus camera and its accompanying software. The acquired images were converted to raw bitmaps and then saved to portable network graphics (PNG) format using lossless compression. The raw bitmaps contained nothing but the pixel data which guaranteed the removal of hidden metadata. The procedures were applied since it was unknown whether the images contained hidden information or not. Otherwise if the images are secured and free of patient information the accompanying information, such as metadata, can be used, for example, to store camera parameters used in the imaging.

3.4.2 Image annotations as the ground truth

In general, the image annotations are essential for training supervised algorithms as well as for their evaluation and comparison. Such information is typically collected by manually annotating a set of images. In face recognition, for example, a ground truth contains identifiers of persons in the images, and often also the locations of facial landmarks, such as eye centres, which can be very useful in training the methods. Commonly, simple tailored tools are used to collect the data, but also generic applications are available for problems which require an exhaustive amount of image data, e.g., LABELME [139] Web tool for annotating visual object categories. Annotating medical images is not an exception, but the two essential considerations apply: 1) annotations must be performed by

clinically qualified personae (specialised or specialising medical doctors, or other trained professionals for specific tasks), denoted as “experts”, and 2) the ground truth should include annotations from several experts.

A more technical problem is to develop a reusable tool for the annotation task. To avoid biasing the results, the experts should be given minimal guidance for their annotation work. Basic image manipulation, such as zoom and brightness control, for viewing the images is needed, and a set of geometric primitives is provided for making the spatial annotations. In LABELME [139], the only primitive is polygon region defined by an ordered set of points. A polygon can represent an arbitrarily complex spatial structure, but ophthalmologists found also the following primitives useful: small circle, which can be quickly put on a small lesion, and circle area and ellipse area which are described by their centroid, radius/radii, and orientation (ellipse). The system also requires at least one representative point for each lesion. This point should represent the most salient cue, such as colour or texture, that describes the specific lesion. Furthermore, a confidence from the set of three discrete values, low, moderate or high, is required for every annotation. The experts are allowed to freely define the types of annotations, i.e., the class labels for the lesions, but typically it is preferable to agree the labels beforehand (e.g., in DIARETDB1: Hard exudates, Soft exudates, Microaneurysms, Haemorrhages). An important design choice is related to the usability of the tool with respect to its graphical user interface (GUI). For example, the GUI should not use colours which distract the annotators from image content.

Development of an annotation tool may take undesirable amount of research time and resources. To help other researchers in this task the tool is made publicly available at [6] as Matlab M-files and as a Windows executable. Matlab is not the optimal environment for developing GUI based applications, but it is widely used in scientific computing, and therefore, familiar to many researchers. Users have full access to the source code which enables tailoring of the tool for their specific needs. The default GUI is shown in Fig. 3.1.

3.4.3 Data format for medical annotations

To store the annotations and to be able to restore their graphical layout, a definition for the data format is required. The data is naturally structured, and therefore structural data description languages are preferred. Several protocols for describing medical data exist, such as HL7 based on the extensible markup language (XML) [1], but these are complex protocols designed for patient information exchange between organisations and information systems. Since the requirements for benchmarking databases in general are considerably less comprehensive, a light-weight data format based on the XML data description language is adopted. Instead of the XML Schema document description, a more compact, and consequently more interpretable Document Type Definition (DTD) description is applied. The used format is given in Listing 3.1.

Listing 3.1: DTD definition

```

<!ELEMENT imgannotooldata (header,markinglist)>
<!ELEMENT header (creator,software?,
                  affiliation?,copyrightnotice)>
<!ELEMENT creator (#PCDATA)>
<!ELEMENT software (#PCDATA)>
<!ATTLIST software version CDATA #REQUIRED>
<!ELEMENT affiliation (#PCDATA)>
<!ELEMENT copyrightnotice (#PCDATA)>
<!ELEMENT imagename (#PCDATA)>
<!ELEMENT imagesize (width,height)>
<!ELEMENT width (#PCDATA)>
<!ELEMENT height (#PCDATA)>
<!ELEMENT markinglist (marking*)>
<!ELEMENT marking ((polygonregion |
                   circleregion | ellipseregion),
                  representativepoint+,confidencelevel,markingtype)>
<!ELEMENT centroid (coords2d)>
<!ELEMENT polygonregion (centroid,coords2d,
                        coords2d,coords2d+)>
<!ELEMENT circleregion (centroid,radius)>
<!ELEMENT ellipseregion (centroid,radius,radius,rotangle)>
<!ELEMENT representativepoint (coords2d)>
<!ELEMENT coords2d (#PCDATA)>
<!ELEMENT radius (#PCDATA)>
<!ATTLIST radius direction CDATA #REQUIRED>
<!ELEMENT rotangle (#PCDATA)>
<!ELEMENT markingtype (#PCDATA)>
<!ELEMENT confidencelevel (#PCDATA)>]>

```

diagnostic test is binary (e.g., normal or abnormal) there are four possible results for the test: true positive, true negative, false positive and false negative. For abnormal test subject, the result is true positive if the diagnostic test outcome is abnormal and false negative if the diagnostic test outcome is normal. For normal test subject, the result is true negative if the diagnostic test outcome is normal and false positive if the diagnostic test outcome is abnormal. An example result of the hypothetical diagnostic test is illustrated in Table 3.3.

Table 3.3: Example diagnostic test result for n test subjects, where TP is the number true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

		True clinical state		Total
		Abnormal	Normal	
Diagnostic test outcome	Abnormal	True positive (TP)	False positive (FP)	Test abnormalities = TP + FP
	Normal	False negative (FN)	True negative (TN)	Test normals = TN + FN
Total		Abnormal subjects = TP + FN	Normal subjects = FP + TN	TP + FN + FP + TN = N

For a given set of subjects, the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) can be used to determine the accuracy of the diagnostic test in form of sensitivity (SN) and specificity (SP) [39]:

$$SN = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{TN + FP}, \quad (3.1)$$

It follows that the sensitivity corresponds to the proportion of correctly diagnosed abnormal test subjects among all abnormal test subjects, and the specificity corresponds to the proportion of correctly diagnosed normal test subjects among all normal test subjects. The sensitivity and specificity are defined in the range [0%, 100%] or [0, 1], and are the fundamental measures of diagnostic accuracy. The sensitivity and specificity evaluate the discrimination power of diagnostic test among abnormal and normal test subjects, but to determine the probability that diagnosed test subject is truly abnormal (or normal) the predictive values can be used:

$$PPV = \frac{TP}{TP + FP}, \quad NPV = \frac{TN}{TN + FN}. \quad (3.2)$$

where the positive predictive value (PPV) indicates the probability that a test subject diagnosed as abnormal by the diagnostic test is really abnormal. Vice versa, the negative predictive value (NPV) indicates the probability that a test subject diagnosed as normal by the diagnostic test is really normal.

More typically the outcome of the diagnostic test is not binary, but a continuous measure or score, such as blood pressure, that indicates the severity or risk of abnormal outcome. In that case, an abnormal diagnostic test can be defined by classifying the test subjects with the score value above the threshold as abnormal, and test subjects with the score value less or equal to the threshold as normal. In other words, the outcome of the diagnostic test is dichotomised using the threshold to produce a binary label (i.e. normal or abnormal). Now, the number of true positives, true negatives, false positives and false negatives can be determined by comparing the binary labels with the true clinical state. For the given threshold, the performance of the diagnostic test can be evaluated, as in the binary classification case, using the sensitivity and specificity.

Let the continuous score values of the hypothetical diagnostic test for given n test subjects be $\zeta = \{\zeta_1, \dots, \zeta_n\}$ and the corresponding true clinical states be $\omega = \{\omega_1, \dots, \omega_n\}$, where the possible clinical state is either normal or abnormal. Then, the probability densities of the diagnostic test produced score values for abnormal and normal test subjects can be denoted as $p(\zeta, abnormal)$ and $p(\zeta, normal)$. If the score values ζ are dichotomised by defining an arbitrary threshold, then it follows that the sensitivity is in fact the proportion of $p(\zeta, abnormal)$ above the threshold, and the specificity is the proportion of $p(\zeta, normal)$ below the threshold. By varying the threshold over the whole range of score values, the sensitivity and specificity vary accordingly generating the receiver operating characteristics curve. The x-axis of a ROC curve is 1-specificity, whereas the y-axis represents directly the sensitivity. Therefore, the ROC curve depicts the trade-off between the sensitivity and specificity and suits well for medical benchmarking. The selection of an operating point from the curve, i.e., fixing either the specificity or sensitivity, is a purely clinical choice. The ROC generation for the hypothetical diagnostic test is illustrated in Fig. 3.2.

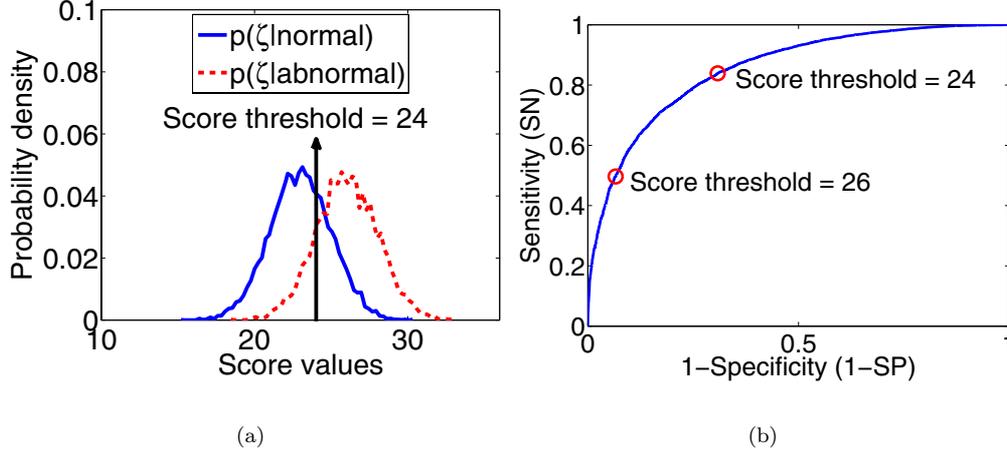


Figure 3.2: Performance evaluation of the hypothetical diagnostic test: a) probability densities of the diagnostic test returned image score values plotted for both populations: normal and abnormal. The area under the normal distribution below the arbitrary threshold represents the specificity, whereas the the area under the abnormal distribution above the same threshold represents the sensitivity; b) by varying the threshold the sensitivity and specificity vary accordingly generating the ROC curve.

3.5.2 ROC-based quality measures

The ROC curve is a solid evaluation method to compare two or more approaches. However, for a large number of curves and when no prior information of an operating point exists, it is useful to form a ROC-based quality measure. A single measure is not as reliable as the full ROC curve, but it enables rough performance comparison. The most simple approach is to fix either sensitivity or specificity and let the other be the measure of performance. For example, diagnosis test A and B result the following (sensitivity, specificity)-pair for fixed sensitivity 0.70 and 0.75:

$$\begin{aligned} A : (0.70, 0.90) \quad \text{and} \quad B : (0.70, 0.85) \\ A : (0.75, 0.87) \quad \text{and} \quad B : (0.75, 0.20). \end{aligned} \quad (3.3)$$

This measure, however, requires the user to define the operational point which can have a dramatic impact on the test results as show in Eq. 3.3. Another straightforward measure would be the total area under the ROC curve (AUC) [70, 22, 69] that can be expressed as an average sensitivity over specificity range or vice versa (Fig. 3.3(a)):

$$AUC = \int_0^1 SN dSP. \quad (3.4)$$

The AUC operates between 0.5 and 1 providing the value of 1 for a perfect diagnostic test and 0.5 for the one that gives random diagnoses. In general, the AUC provides a single summary value for method comparison which can be derived in objective manner without the requirement of user defined parameters. For example, if a ROC curve of diagnostic test A uniformly dominates the ROC curve of diagnostic test B ($SN_A \geq SN_B, \forall SP$) then the AUC correctly ranks the tests in terms of performance ($AUC_A \geq AUC_B$). The integration over the sensitivity values, however, can lead to misleading results when the ROC curves cross. For example, the ROC curve of the diagnostic test A may have higher AUC even though the diagnostic test B show better performance almost over the entire range of the specificities (or alternative the score thresholds). Moreover, the crossing is likely to happen near the points (0,0) and (1,1), where the ROC curves coincide. This is not a desired property if the relevant information lies in the specific part of the ROC curve, e.g. the diagnostic test B should be better due to superior performance in the high specificity range. To overcome this property, an alternative version of AUC was proposed, where the relevant information is solely regarded. The measure is known as partial AUC (PAUC) [30] and can be defined as follows:

$$PAUC = \int_a^b SN dSP, \quad (3.5)$$

where the interval $[a, b]$ within $[0, 1]$ defines the specificity range of particular importance. The direct implication of defining an interval for the integration is that the non-parametric nature of the AUC is lost. Moreover, the PAUC does not actually remove the problem of crossing ROC curves, but restricts the range where the crossings can affect the summary measure. The PAUC is illustrated in Fig. 3.3(b).

A point on ROC curve, where sensitivity is equal to specificity ($SN = SP$), is another indicator of performance used in method comparison [128]. In terms of correct diagnoses, the indicator is known as symmetry point (SYM) while in terms of incorrect diagnoses it is referred as the equal error rate (EER). These measures can be defined as follows when ($SN = SP$):

$$SYM = SN = SP = 1 - EER \quad (3.6)$$

Since SYM and EER are the two sides of the same performance indicator, the following discussion will concentrate on EER. The EER lies between 0 and 0.5 returning value 0 for a perfect diagnostic test and 0.5 for the diagnostic test that returns random diagnoses. Graphically, the EER is defined by the intersection of a ROC curve and the diagonal line (equal error line) drawn from the upper left corner of the ROC plot to the bottom right corner (Fig. 3.3(c)). The rationale behind the EER is that it assumes two balancing misclassification errors in the ROC curve, i.e 1-SN and 1-SP, equally harmful. If the misclassification costs are unknown then this is a reasonable assumption. Since the EER can be observed in terms sensitivity and specificity unlike AUC and PAUC, and it does not require any additional parameters from the user, the EER was considered suitable for this study, although it may not be the best choice for the screening purposes. The EER values were computed from the nearest actual points on the ROC curve without interpolation to derive the true lower bound for the error. However, if prior information

of risks or penalties exists, and the cost ratio can be fixed then the minimum weighted error rate can be used, where the sensitivity (SN) and specificity (SP) are combined as

$$WER(\hat{R}) = \frac{FPR + \hat{R} \cdot FNR}{1 + \hat{R}} = \frac{(1 - SP) + \hat{R} \cdot (1 - SN)}{1 + \hat{R}}, \quad (3.7)$$

where $\hat{R} = C_{FNR}/C_{FPR}$ is the cost ratio between false negative rate (FNR) and false positive rate (FPR). The WER was adopted from ROC related detection error trade-off (DET) curve used in face verification [16]. A ROC curve with WER is shown in Fig. 3.3(d).

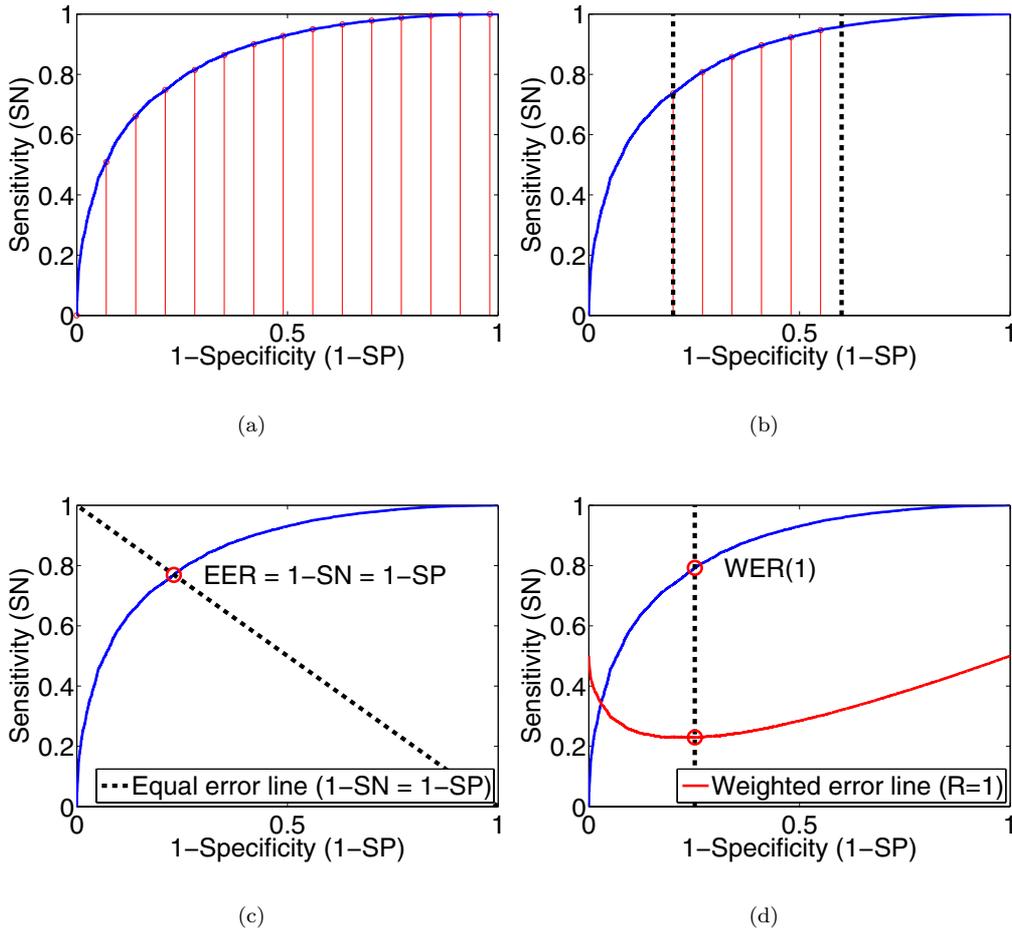


Figure 3.3: Graphical illustrations of ROC-based quality measures: a) area under the ROC curve; b) partial area under the ROC curve within specificity interval [0.4 0.8] c); Equal error rate; d) weighted error rate ($\hat{R} = 1$).

3.5.3 Image-based evaluation

The automatic image based evaluation follow the medical practise, where the decisions are “subject-wise”. The image analysis system is treated as a black-box which takes an image as the input. If the images are assumed to be either normal or abnormal, then the system produces a score that corresponds to the probability of the image being abnormal, and a high score corresponds with high probability. The objective of the image-based evaluation protocol is to generate a ROC curve by manipulating the score values of the test images. The practises were adopted from [36].

Let the image analysis algorithm produced score values for n test images be $\zeta^{im} = \{\zeta_1^{im}, \dots, \zeta_n^{im}\}$ and the corresponding image-wise ground truths be $\omega^{im} = \{\omega_1^{im}, \dots, \omega_n^{im}\}$, where the ω^{im} is either “normal” or “abnormal”. Then, by selecting a threshold for the score values (ζ^{im}), the test images can be classified either normal or abnormal, and the performance expressed in the form of sensitivity and specificity can be determined by comparing the outcome with the corresponding image-wise ground truth (ω^{im}). If the same procedure is repeated using each test image score as a threshold, the ROC curve can be automatically determined since each threshold generates a (sensitivity, specificity)-pair that is a point on the ROC curve. Consequently, the procedure requires that the test images include samples from both populations, normal and abnormal. The image score based evaluation method is presented in Algorithm 1.

Algorithm 1 Image-wise evaluation based on image scores

```

1: for each test image do
2:    $TN \leftarrow 0, TP \leftarrow 0, FN \leftarrow 0, FP \leftarrow 0$ 
3:   curr_score  $\leftarrow$  image score
4:   for each test image do
5:     if curr_score  $\geq$  image score then
6:       if ground truth assignment = “normal” then
7:          $TN = TN + 1$ 
8:       else
9:          $FN = FN + 1$ 
10:      end if
11:     else
12:       if ground truth assignment = “abnormal” then
13:          $TP = TP + 1$ 
14:       else
15:          $FP = FP + 1$ 
16:       end if
17:     end if
18:   end for
19:    $SN = \frac{TP}{TP + FN}$  (Sensitivity)
20:    $SP = \frac{TN}{TN + FP}$  (Specificity)
21:   Add new ROC point  $(x; y) = (1-SP, SN)$ 
22: end for
23: Return the final ROC curve (all points)

```

3.5.4 Pixel-based evaluation

To validate a design choice in method development, it may be useful to measure also the spatial accuracy, i.e., whether the detected lesions are found in correct locations. Therefore, a pixel-based evaluation protocol which is analogous to the image-based evaluation is proposed. In this case, the image analysis system takes an image as the input and outputs a likelihood score for each pixel. The objective of the pixel-based evaluation is to generate a ROC curve which describes the pixel-level success.

Let the image analysis algorithm produced pixel score values for all n pixels in test set be $\zeta^{pix} = \{\zeta_1^{pix}, \dots, \zeta_n^{pix}\}$ and the corresponding pixel-wise ground truth be $\omega^{pix} = \{\omega_1^{pix}, \dots, \omega_n^{pix}\}$, where the ω^{pix} is either “normal” or “abnormal”. Then, by selecting a global pixel-wise threshold for the pixel score values (ζ^{pix}), the pixels in all images can be classified to either normal or abnormal. Now, the sensitivity and specificity can be computed by comparing the outcome to the pixel-wise ground truth (ω^{pix}). If the procedure is repeated using each unique pixel score as threshold, the ROC curve can be automatically determined. The pixel-wise evaluation procedure is given in Algorithm 2. Note that the abnormal test image pixels contribute to both sensitivity and specificity, whereas the normal images only contribute to the specificity.

Algorithm 2 Pixel-wise evaluation based on pixel scores

```

1: Form a list of tested pixel scores
2: for each tested pixel score (curr_pix_score) do
3:   TN ← 0, TP ← 0, FN ← 0, FP ← 0
4:   for each test image do
5:     for each test image pixel score do
6:       if curr_pix_score ≥ pixel score then
7:         if ground truth pixel assignment = “normal” then
8:           TN = TN + 1
9:         else
10:          FN = FN + 1
11:        end if
12:       else
13:         if ground truth pixel assignment = “abnormal” then
14:           TP = TP + 1
15:         else
16:           FP = FP + 1
17:         end if
18:       end if
19:     end for
20:   end for
21:    $SN = \frac{TP}{TP + FN}$  (Sensitivity)
22:    $SP = \frac{TN}{TN + FP}$  (Specificity)
23:   Add new ROC point (x; y) = (1-SP, SN)
24: end for
25: Return the final ROC curve (all points)

```

The evaluation forms a list of global pixel-wise scores from the test image pixel scores which determines the score thresholds. The use of all unique pixel scores in the test images is time-consuming if the number of images in the test set is large or high resolution images are used. The problem can be overcome by sampling the test image pixel scores. To preserve the test set’s pixel score distribution, the global threshold scores can be devised as follows: 1) sort all the unique pixel scores in ascending order to form an ordered sequence L , and 2) compose the new reduced sequence of pixel scores $L_{sampled}$ by selecting every j th likelihood in L .

3.6 Baseline algorithm

Images, ground truth, and the evaluation protocol are sufficient to establish a public benchmarking database. However, by defining the baseline results or even providing the baseline algorithm, the users may find it easier to evaluate the maturity and applicability of their methods by themselves. A strawman algorithm that has been commonly agreed as benchmark for the method comparison within the research field would be the best choice for the purpose [163]. The baseline method for DIARETDB1, however, is selected purely based on practical reasons, and therefore it does not fulfil the characteristics of the strawman algorithm. Regardless, it can be used to set the baseline results for the DIARETDB1 database. The goal of introducing a new baseline algorithm for DIARETDB1 will be discussed after the method description.

The baseline algorithm tackles the detection problem as two-class classification problem, where the possible classes are “lesion type present” and “lesion type absent”. Thus, the images or pixels either contain a certain type of lesion or not. Instead of modelling both classes, “lesion type present” and “lesion type absent”, the baseline algorithm distinguishes a lesion type from all other possible objects in eye fundus images by modelling only that certain lesion type. This is known as the one-class classification scheme or novelty detection. Because the DIARETDB1 contains four types of lesions, i.e haemorrhages (HA), microaneurysms (MA), hard and soft exudates (HE and SE), the baseline algorithm tackles four two-class prediction problems in total. The set of DIARETDB1 lesion types is denoted with $\xi = \{HA, MA, HE, SE\}$

The lesions are modelled based on their colour information which is the information driving the analysis of diabetic lesions in medical practice as well. In the modelling, the class-conditional probability density function of certain lesion type with respect to colour pixel value is estimated as a weighted sum of Gaussians. The Gaussian mixture model probability density function can be defined as

$$p(\mathbf{x}|\xi_i) = p(\mathbf{x}; \boldsymbol{\theta}_{\xi_i}) = \sum_{c=1}^C \alpha_c \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c), \quad (3.8)$$

where \mathbf{x} is the colour pixel value (e.g. RGB), ξ_i is the i th lesion type in ξ and α_c is the weight of the c th component. The weight can be considered as a prior probability that a value of the random variable is generated by the c th component, and thus $0 < \alpha_c < 1$ and $\sum_{c=1}^C \alpha_c = 1$. The c th component is represented as multivariate normal distribution

$N(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ defined by the mean $\boldsymbol{\mu}_c$ and covariance $\boldsymbol{\Sigma}_c$. It follows that the Gaussian mixture model probability density function is defined by a parameter list

$$\boldsymbol{\theta}_{\xi_i} = \{\alpha_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \dots, \alpha_C, \boldsymbol{\mu}_C, \boldsymbol{\Sigma}_C\}. \quad (3.9)$$

To estimate the model parameters, a completely automatic Gaussian mixture model (GMM) estimation is performed by using the Figueiredo-Jain (GMM-FJ) algorithm [41] for which a public implementation is available [124]. The estimation is conducted separately for each lesion type in $\boldsymbol{\xi}$. To determine if lesion type is present, the baseline algorithm uses thresholding as follows:

$$\begin{aligned} p(\mathbf{x}|\xi_i) &> t_{\xi_i}^{pix} \\ p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n|\xi_i) &> t_{\xi_i}^{im}, \end{aligned} \quad (3.10)$$

where the lesion type ξ_i is present in given image pixel if the pixel-likelihood $p(\mathbf{x}|\xi_i)$ is greater than the decision threshold $t_{\xi_i}^{pix}$, and in given image if the image-likelihood $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n|\xi_i)$ is greater than the decision threshold $t_{\xi_i}^{im}$. The image-likelihood $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n|\xi_i)$ is derived from the pixel-likelihoods and will be further discussed in Chapter 4.

In the training of the baseline algorithm, the colour pixel values for each lesion type are extracted from the ground truth regions of the DIARETDB1 training images. Thus, the training data comprise all the extracted colour pixel values and a label for each colour pixel value denoting the lesion type. The GMM-FJ estimation is then applied to the training data to estimate the model parameters of class-conditional probability density functions. Since the GMM-FJ algorithm adjusts the number of components during the estimation by annihilating the components that do not support the data, the estimation is started with relatively large number of components. The initial guesses for component means are distributed uniformly into the whole space occupied by training samples, whereas the component covariances were identical and selected over the whole data. For all the lesions, the initial number of component was fixed to ten components which is considered feasible, since the estimated probability density functions had always less components in the experiments. Once the training of baseline algorithm is conducted, a Gaussian mixture model probability density function for each lesion type ξ_i is returned.

In the testing of the baseline algorithm, a likelihood map for each test image in the DIARETDB1 database is computed according to the estimated class-conditional probability density functions. The DIARETDB1 database comprise four lesion types, and thus four likelihood maps are computed. The pixel value in each likelihood map denotes the likelihood that it contains the corresponding lesion type.

In the evaluation of the baseline algorithm, the detection performance is evaluated using the methodology described in Section 3.5. The pixel and image scores required for the evaluation are defined as follows:

$$\begin{aligned} \zeta_{\xi_i}^{pix} &= p(\mathbf{x}|\xi_i), \\ \zeta_{\xi_i}^{im} &= p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n|\xi_i), \end{aligned} \quad (3.11)$$

where $\zeta_{\xi_i}^{pix}$ and $\zeta_{\xi_i}^{img}$ are the pixel and image score values. Thus, the evaluation in the case of baseline algorithm finds for each lesion type the likelihood thresholds $t_{\xi_i}^{im}$ and $t_{\xi_i}^{pix}$ that best discriminates the lesioned images and pixels from non-lesioned ones with respect to the sensitivity and specificity. It should be noted, however, that the evaluation protocols were designed not take into consideration on how the score values are determined. Before the baseline algorithm can be actually applied there are two important problems to be solved: 1) how to combine the medical annotations of the DIARETDB1 database collected from several medical experts into unambiguous ground truth to train the baseline algorithm, and 2) how to form an image-wise score $p(\mathbf{x}_1, \mathbf{x}_2 \dots, \mathbf{x}_n | \xi_i)$ from the result of the baseline algorithm (likelihood maps) to perform the subject-wise (image-based) evaluation. In Chapter 4, these two problems are investigated in detail and good strategies are proposed for the baseline algorithm.

The baseline algorithm is not intended as the state-of-the-art algorithm, but as a statistical tool to study the colour properties of eye fundus images with diabetic retinopathy, where an important property is that it can be applied in detection of any lesion type. The probability density estimation and the resulting likelihood maps provide the necessary means for such statistical investigation, and therefore the baseline algorithm fulfils the requirements it was designed for. Since the baseline algorithm was required for the research in any case, it was only practical to set the baseline results for the DIARETDB1 database using the baseline algorithm. As a result, the functionality of the baseline algorithm and the benchmarking framework were verified simultaneously. Otherwise, the benchmarking framework would have required a baseline method from the existing work for each lesion type separately, since the state-of-the-art lesion detection methods are typically designed for certain lesion type.

Another important property of baseline algorithm is obtained from the one-class classification scheme. Since a lesion is modelled by estimating the density of the training colour pixel values, no assumptions are made about other retinal structures including other lesion types. Thus, high likelihoods are assigned to pixels that resemble the estimated density functions including the incorrect non-lesion pixels. This is an important property, since it reveals the colour dependencies in the eye fundus image exclusively with respect to the modelled lesion type. It also indicates the goodness of colour in the detection of diabetic lesions.

3.7 Case studies – DIARETDB databases

During the course of work two eye fundus image databases with ground truth were published: DIARETDB0 and DIARETDB1. The work on the DIARETDB0 database provided the essential information on how to collect, store, annotate, and distribute the eye fundus images, and therefore the DIARETDB1 database was a natural continuation to establish a better database for algorithm evaluation. A full description of the databases follows including a description of non-public eye fundus image database (BRISTOLDB) involved in the experimental work of this thesis. Furthermore, the database review in Section 3.3 comprised two diabetic retinopathy databases for benchmarking, i.e. ROC and MESSIDOR, of which the ROC database is evaluated against the DIARETDB1 database in this section. The MESSIDOR database did not fulfil the minimal benchmarking requirements, and therefore the main focus is put on the ROC database.

3.7.1 DIARETDB0

The database consists of 130 color eye fundus images of which 22 are normal and 108 contain signs of the diabetic retinopathy (microaneurysms, haemorrhages, hard and soft exudates, and neovascularisation). The images were taken by the Kuopio university hospital with multiple 50 degree field-of-view digital eye fundus cameras. The images were dedicatedly selected, but their distribution does not correspond to any typical population, i.e., the data is biased – no prior information can be and should be devised from it. The amount of imaging noise, projective camera geometry, and optical aberrations (dispersion, transverse and lateral chromatic, spherical, field curvature, coma, astigmatism, distortion) are unknown due to unknown camera settings, and thus the variance over the visual appearance of the diabetic lesion types can be considered as maximal. The images are in PNG format from where metadata is explicitly removed by converting the images into raw RGB data and then regenerating the PNG images. Each eye fundus image in the database is accompanied by a text file that lists the diabetic lesion types in the image as follows

```
redsmalldots hemorrhages hardexudates softexudates neovascularisation
```

The DIARETDB0 comprise nine pre-defined train and test sets which were selected as follows. Firstly, the images were partitioned into five groups in such a way that if the groups were split in two the diabetic lesion types and normal images would be evenly distributed between the sets. Secondly, the train and test sets were prepared by first selecting n random images from each group to the train set and then placing the remaining images to the test set. The nine different train and test sets were generated by varying the value of n : 2, 4, 6, 8, 10, 12, 14, 16 and 18. The characteristics of the train and test sets are given in Table A.1 in Appendix A. The images, ground truth and the predefined train and test sets, can be downloaded from [3].

It should be noted that the DIARETDB0 database was not used in the experimental work of this thesis for the following reasons. The DIARETDB1 database was considered as a good starting point for experimental work, since the images were taken with single eye fundus camera. Thus, the images contained the same optical aberrations and the photometric content over the images was more consistent than if multiple cameras were used as in the DIARETDB0 database. A natural continuation for the experimental work would have been DIARETDB0 database, but unfortunately the lack of spatial groundtruth restricted the use of baseline algorithm and the BRISTOLDB was used instead.

3.7.2 DIARETDB1

The database consists of 89 colour fundus images, of which 84 contain at least mild non-proliferative signs of diabetic retinopathy and 4 images do not contain any signs of diabetic retinopathy according to all experts participated in the annotation. The images were taken by the Kuopio University Hospital with single 50 degree field-of-view digital eye fundus camera¹. The images were dedicatedly selected by the medical experts, but

¹ZEISS FF 450^{plus} fundus camera with Nikon F5 digital camera

their distribution does not correspond to any typical population, i.e., the data is biased – no a priori information can or should be devised from it. The amount of imaging noise is unknown, but the images share the same optical aberrations (dispersion, transverse and lateral chromatic, spherical, field curvature, coma, astigmatism, distortion), and thus the variance over the visual appearance of the diabetic lesion types can be considered smaller than in the DIARETDB0 database. The diabetic abnormalities in the DIARETDB1 images are relatively small, but since they appear near the macula they are considered to threaten the eyesight. The images are in PNG format from where metadata is removed similarly as in DIARETDB0 by converting the images into raw RGB data and then regenerating the PNG images. Each eye fundus image in the database is accompanied by a XML-file that contains the medical expert provided annotations.

Pixel-wise annotations were collected independently from four medical experts by providing an image annotation software described in Section 3.4.2. The medical experts were asked to annotate image regions related to microaneurysms, haemorrhages, hard and soft exudates, report their confidence, and annotate the single most representative point within each image region. The experts were further instructed to enclose the lesion within the annotation. The confidence denoted the expert’s subjective confidence that the annotated region was the given lesion type, where the used confidence scale, $\{< 50\%, > 50\%, 100\%\}$, was collectively agreed by the medical experts. It should be noted that the uninstructed annotation process caused notable differences between the medical experts. The discussion on how to devise the ground truth for the DIARETDB1 database from the medical expert annotations is given in Section 4.3.

Train and test set for the DIARETDB1 database were selected as follows. Firstly, the images were partitioned into five groups in such a way that if the groups were split in two the diabetic lesion types and normal images would be evenly distributed between the sets. Secondly, the images within each group were sorted into descending order according to the number of pixels annotated in the image. Two similar image sets were formed then by distributing every second sorted image in each group to the train set and placing the remaining images to the test set. Lastly, the images without annotations were moved from the train set to the test set by hand. The characteristics of the train and test set are given in Tables A.2 and A.3 in Appendix A and the files for the corresponding image sets are listed in Tables A.4 and A.5. The experimental work of this thesis is mainly based on the DIARETDB1 database and for that reason the train and test files were also listed in this thesis. The images, expert annotations and the pre-defined train and test sets can be downloaded from [4]. The web site will also provide a development kit for the DIARETDB1 database containing full Matlab functionality (M-files) for reading and writing images and the expert annotations. The annotation tool will be available at [6].

DIARETDB1 VERSUS ROC

In the ROC database, the image analysis algorithm outputs the microaneurysm location, likelihood that it was a microaneurysms and the index of the image. This is then compared to the ground truth of ROC database, where an image location and radius represents a microaneurysm. If the location of the microaneurysm detected by the image analysis algorithm falls within the radius of the closest ground truth microaneurysm, the detection is considered correct and otherwise it is a false detection. The FROC curve

is generated by manipulating the likelihoods of the detected microaneurysms. In the proposed DIARETDB1 database, the evaluation problem is tackled by evaluating the detection success on both pixel- and image-level, where the ground truth is annotated with pixel accuracy and the image-level ground truth is inferred from the pixel-level ground truth. In total, the DIARETDB1 comprise ground truth for four lesion types, i.e. haemorrhages, microaneurysms, hard and soft exudates. The image analysis algorithm outputs a score value for each pixel and image, where the pixel-score denotes the probability that a certain lesion type is present in the pixel and the image-score denotes the probability that a certain lesion type is present in the image. If the high score values are assigned to correct pixels and images that contain the lesion type in ground truth, the evaluation protocol returns high performance. By manipulating the pixel and image score values separately, a ROC curve for both image and pixel-level detection success is generated.

The ground truth for the ROC database was collected from multiple medical experts by annotating the location, radius for each microaneurysm. The microaneurysms for which two experts assigned the label “microaneurysm” was selected to be the ground truth. The ground truth microaneurysms were further processed by asking an expert to assign his/hers subjective estimate for contrast and visibility, where the possible choices were subtle, regular or obvious. In the DIARETDB1 database, the ground truth was also collected from multiple medical experts by annotating the region enclosing the lesion, confidence and the single most representative point within the region. The confidence denoted the expert’s subjective confidence that the annotated region was the given lesion type. The final ground truth image for each lesion type in DIARETDB1 database was devised by thresholding the mean expert confidence computed for each pixel in the image. The mean expert confidence was computed over the experts that had made annotations in the image for that specific lesion type.

The above mentioned evaluation choices made for ROC database are well-reasoned for microaneurysm detection, where the lesion is small and round shaped. In the case of other diabetic lesion types, however, the choices may lead to unwanted results. A part of the problem seems to arise from the difficulty to find a general parametric representation of a lesion that could be applied in benchmarking of any type of lesion. A single image point and radius do not represent well lesions which are not round shaped, but have irregular outline and large scale variation, e.g., hard exudates. In addition, the lesion-based performance evaluation does not actually reflect the detection success of diabetic retinopathy, even though it is probably the most important objective in microaneurysm detection. It would be also informative to know how well the image analysis algorithm segments the lesions instead of just knowing the trade-off between the proportion of correctly detected lesions and the average number of false detections. These generalisation problems to other lesion types may lead to the development of lesion type specific evaluation protocols and further complicate the performance evaluation if the ROC database is further extended.

In comparison to the ROC database, the evaluation in DIARETDB1 database is conducted on pixel and image basis, which does not require parametric definition for lesions, but a ground truth that denotes the lesioned pixels and images. Thus, the evaluation protocol can be applied to any type of lesion using the same framework, where the pixel-level evaluation allows to measure how well the segmentation is conducted. In addition,

the two-level performance evaluation allows to connect the pixel-level detection problem to its objective, e.g. by assessing the lesion segmentation performance on pixel-level and the performance of diabetic retinopathy detection on image-level. Unfortunately, the DIARETDB1 database does not currently contain ground truth for the diabetic retinopathy, and therefore both the pixel- and image-level evaluations concern lesion detection. The pixel-level accuracy in evaluation also induce other problems, such as how the experts should be instructed to achieve optimal annotation accuracy without introducing bias. In particular due to the annotation process, the current spatial ground truth of the DIARETDB1 database is not optimal for the pixel-level performance evaluation.

The ground truth collection for both databases is surprisingly similar. Manual annotations were collected from multiple experts from where the final ground truth was then devised based on the expert agreement. In addition, each annotation was assigned with a confidence that denoted expert's subjective certainty. In the ROC database, however, the confidence was collected from single expert and assigned after selecting the ground truth, whereas in the DIARETDB1 database the confidence was gathered during the annotations before devising the ground truth. Thus, all the experts affect with their subjective confidence on how the ground truth is generated unlike in the ROC database. Another issue concerns the annotation process, and thus both databases. The ground truth was devised from manual annotations which is laborious and time-consuming to collect, and therefore complicates the construction of large benchmarking databases. In addition, it is difficult to measure the quality of the annotations.

The usability of the ROC database, however, is already at more mature level than DIARETDB1 database. The fact that the ground truth for the test set is withheld unknown and the evaluation is conducted online via web-based interface, the cheating is efficiently prevented. Thus, the algorithm comparison can be regarded as fair. These shortcomings in DIARETDB1, however, point out clear steps for the future work.

3.7.3 BRISTOLDB

BRISTOLDB is a non-public eye fundus image database that was collected by the Bristol Eye Hospital for the development of automated (hard) exudate identification system [119]. The database comprises a total of 142 eye fundus images of which 107 images (17 normal and 90 abnormal images) were received from the author of the database and used in the experiments. The images were captured by an ophthalmologist using a standard 45 degree field-of-view digital eye fundus camera (Canon CR6-45 Non-Mydriatic, CR6-45NM) with a modified digital backunit (Sony PowerHAD 3CCD colour video camera and Canon CR-TA). The hard exudate pixels, the BRISTOLDB ground truth, in the images were manually annotated by a consultant ophthalmologist (Fig. 3.4). It is worth noting that patients who had undergone the laser treatment were removed from the database while collecting the images. The BRISTOLDB data is illustrated in Fig. 3.4. The train and test set for BRISTOLDB was selected by choosing the first 30 images with hard exudates to the train set and then placing the remaining 77 images to the test set. The characteristics of the train and test sets are given in Table A.6 in Appendix A.

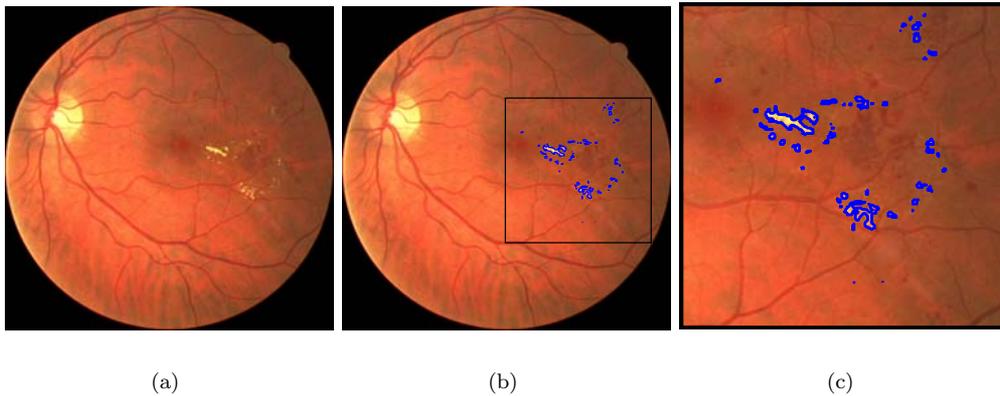


Figure 3.4: Example image in the BRISTOLDB database and the corresponding expert annotations [119]: (a) abnormal image showing hard exudates; (b) expert annotations; (c) close-up of the expert annotations.

3.8 Summary

In this chapter, a set of guidelines were given for constructing benchmarking databases from the scratch for eye fundus image analysis algorithms. The guidelines were based on the demanding, laborious and time consuming experiences in collecting the benchmarking databases DIARETDB0 and DIARETDB1 for diabetic retinopathy detection. As a result, three mandatory components were described that are required in the benchmarking: 1) true patient images, 2) ground truth, and 3) an evaluation protocol. In the collection of the true patient images and ground truth one should most importantly remember the distribution correspondence with the desired population, the privacy protection of the patient data, and the use of several clinically qualified persons in the ground truth collection. For the evaluation, an image and pixel-based evaluation approaches were given that were based on the receiver operating characteristics analysis. In addition, the DIARETDB0 and DIARETDB1 databases, the DIARETDB1 development kit and the image annotation tool are publicly available to promote and help other researches to collect and publish their data. It should be noted that the methodology presented in this chapter is based on ROC-analysis which is widely used in eye fundus image analysis, but it is not perhaps the only possibility for the detection of diabetic retinopathy. The medical image analysis in general could offer alternative options.

Overall image score and multiple expert information

Two problems essential for supervised learning and classification in eye fundus image processing is addressed in this chapter: 1) how to fuse medical annotations collected from several medical experts into unambiguous ground truth and for training a classifier, and 2) how to form an image-wise overall score for accurate and reliable automatic eye fundus image analysis. The first problem arises from the typical need to collect a medical ground truth from several experts. Currently, it is unclear how these expert opinions (annotations) should be combined for the use of image analysis algorithms designed for screening diabetic retinopathy. The second problem is due to the practice to diagnose the diabetic retinopathy subject-wise. The eye fundus image analysis algorithms typically produce spatially distinct results that should be combined into a single image-wise score value for the subject-wise decision making. These two problems are investigated in detail, good strategies are proposed and experimental results on the DIARETDB1 database are reported. Both these problems are addressed by applying the baseline algorithm and the benchmarking framework presented in Chapter 3.

4.1 Introduction

While constructing the benchmarking framework, it became evident that there are certain crucial research problems that need to be studied further. The fusion of annotations from several experts into single unambiguous ground truth, that can be applied in algorithm comparison and training a classifier, was one of them. At least for eye fundus images, this is of particular importance since the opinions of medical experts may deviate from each other or the experts may graphically describe the same finding in very different ways. This can be partly avoided by instructing the doctors while annotating, but as mentioned, often this is not desired since the data can be biased and grounds for understanding the phenomenon may weaken. Therefore, it is necessary to study appropriate fusion or voting methods. It is worth noting that in computer vision, e.g. in visual object categorisation, this problem has not been addressed at all (e.g., the recent LABELME database [139] or the popular CALTECH101 [40]).

Another problem arises from the difference on how the medical experts see the information in eye fundus images in comparison to automatic image analysis algorithms. Instead of searching for lesioned pixels or regions in the image, the medical experts attempt to make diagnosis for the whole image using the whole depicting supporting information in the image. Thus, the problem is how to combine a single overall image score from the output of eye fundus image analysis algorithm that could be used as a diagnostic test measure. Also this problem is less known in eye fundus image processing, but a studied problem within the context of multiple classifiers or classifier ensembles (e.g., [92, 161, 51]). Both of the presented problems are studied using the benchmarking framework and the baseline algorithm covered in Chapter 3 (Fig. 4.1).

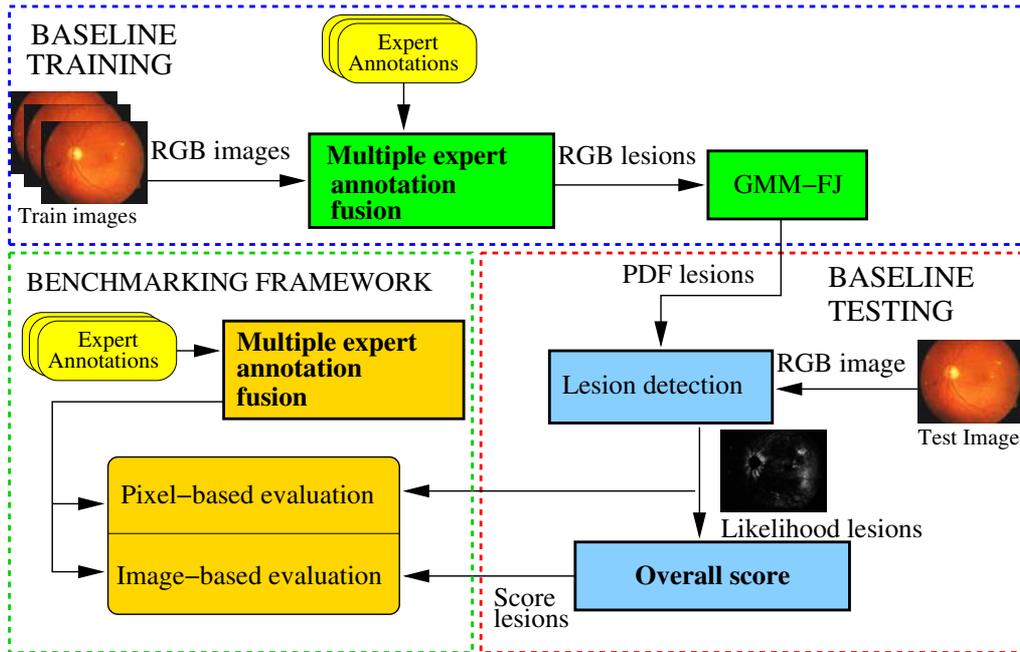


Figure 4.1: Overall score and multiple expert information fusion in the context of the baseline algorithm and benchmarking framework.

The content of this chapter is reported in [89] and the chapter is organised as follows. The presented problems are discussed in detail in Sections 4.2 and 4.3. The proposed fusion and scoring methods are experimentally evaluated using the baseline algorithm and the benchmarking framework presented in Section 4.5. Based on the discussions and the presented empirical results, conclusions are drawn and the best practises for the baseline algorithm are devised.

4.2 Overall image score selection

The medical diagnosis is typically based on prior information and laboratory tests under strong Bayesian decision making [56]. This procedure can be interpreted as a classifier

that takes the medical knowledge concerning the patient as its input and outputs the decision weighted with the prior information (background knowledge). In eye fundus image analysis, the medical knowledge is typically described using image features (colour, texture, intensity, shape, etc.) which are directly extracted from the pixels and blocks of an image. If the medical knowledge extracted from the eye fundus image is composed of n features, $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n$, where each feature is a vector, then the output of an eye fundus image analysis algorithm can be denoted as $p(\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n)$. The joint probability can be approximated from the classification results (likelihoods) in terms of decision rules using ideas from the combined classifier theory (classifier ensembles) [92]. The most plausible combination rules applicable also here are the product, sum (mean), maximum and median rules. The median rule was replaced with a more intuitive rank-order based rule for the case: summax-rule, i.e., the sum of some proportion of the largest values ($summax_X\%$). Now, the decision rules to form an (overall) image score for the baseline algorithm, where the extracted features are colour pixels values, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, can be written as follows:

$$\begin{aligned}
\zeta_{\xi_i}^{im} &= \max_{j \in n} \{p(\mathbf{x}_j | \xi_i)\}, & (\text{Max rule}) \\
\zeta_{\xi_i}^{im} &= \frac{1}{n} \sum_{j=1}^n p(\mathbf{x}_j | \xi_i), & (\text{Mean rule}) \\
\zeta_{\xi_i}^{im} &= \prod_{j=1}^n p(\mathbf{x}_j | \xi_i), & (\text{Product rule}) \\
\zeta_{\xi_i}^{im} &= \sum_{j \in N} p(\mathbf{x}_j | \xi_i), & (\text{Summax rule})
\end{aligned} \tag{4.1}$$

where $\zeta_{\xi_i}^{im}$ is the baseline algorithm produced image score for the lesion type ξ_i and N is the set of $X\%$ of the largest likelihoods $p(\mathbf{x}_j | \xi_i)$. Thus, each colour pixel value \mathbf{x}_j votes that the lesion type ξ_i is present in the image with the confidence $p(\mathbf{x}_j | \xi_i)$. In the formulation, the maximum and sum rules can be seen as two extrema whereas summax operates between them so that X defines the operation point.

4.3 Ground truth estimation from multiple expert annotation

The performance of an eye fundus image analysis algorithm is typically evaluated by comparing the algorithm result to some known reference which is commonly referred as ground truth. In case of lesion detection, where the objective is to segment objects of certain lesion type, the algorithm result and the ground truth are usually represented as binary maps, where the pixel value denotes the binary decision for the object being present in that pixel location. Hence, the ground truth is most commonly collected by manually annotating the objects in the eye fundus image that should be detected. By comparing the algorithm produced segmentation result and the annotated ground truth using some predefined evaluation metric, the detection performance is then evaluated in

terms of similarity or overlap as discussed previously in the benchmarking framework. The manual annotations are unfortunately subjective and opinions of medical experts may deviate, and therefore it is recommended in eye fundus image analysis to collect annotations from multiple experts. Multiple annotations on the other hand generate a new problem of how the annotations should be combined into a single unambiguous ground truth. Currently, this problem has been recognised in the field of eye fundus image analysis, but not comprehensively studied. Fortunately, there are works in the field of medical image analysis that can be considered for the task, such as voting policy, variation minimisation, simultaneous truth and performance level estimation (STAPLE), and truth estimate from self distances (TESD).

Consider an eye fundus image for which n ophthalmologists annotated objects that should be automatically detected. Furthermore, let the binary map \mathbf{D}_i represent the annotation of i th expert, where the pixel value $D_i(x, y) = 1$ indicates the object locations in the eye fundus image. Now, the task is to find such binary map \mathbf{GT} using the expert annotations \mathbf{D}_i that would best describe the true object locations in the image. Four different approaches from the literature for estimating such binary map is described next.

The most straightforward solution to estimate the ground truth is to use the voting policy, where each expert votes with the pixel value $D_i(x, y)$ that the object is present in that corresponding image location [101]. A voting threshold is then used to determine if the location is truly part of the object as follows:

$$GT(x, y) = \begin{cases} 1, & \text{if } \sum_{i=1}^n D_i(x, y) > thr_{vote} \\ 0, & \text{otherwise} \end{cases}, \quad (4.2)$$

where thr_{vote} is the voting threshold and $GT(x, y)$ is the binary pixel value of the ground truth. The advantage of the voting policy is that it is simple and intuitively corresponds to colleague consultation, where each expert is equally important.

The variation minimisation provides an alternative approach for the voting policy, where the ground truth estimation is considered as an optimisation task. The unknown ground truth is found by minimising the overall difference between the ground truth and expert annotations which can be formulated as follows [101]:

$$E = \arg \min_{\mathbf{GT}} \sum_{i=1}^n \sum_{\forall x, y \in \mathbf{D}_i} |GT(x, y) - D_i(x, y)|, \quad (4.3)$$

where E is the energy that should be minimised. Thus, the final ground truth is the binary map \mathbf{GT} that produce the minimum energy.

A more elaborate ground truth estimation approach is proposed by Warfield et. al. [184] which is known as simultaneous truth and performance level estimation (STAPLE). In this method, the unknown ground truth is estimated by maximising the expert annotation performances in comparison to the ground truth, where the annotation performance of each expert is measured with (sensitivity, specificity)-pair. If the annotation performance of i th expert is denoted with (SN_i, SP_i) then the objective is to estimate such vectors

$\mathbf{SN} = [SN_1, SN_2, \dots, SN_n]$ and $\mathbf{SP} = [SP_1, SP_2, \dots, SP_n]$ characterising the annotation performances of n experts that maximise the following complete log-likelihood function:

$$(\hat{\mathbf{SN}}, \hat{\mathbf{SP}}) = \arg \max_{\mathbf{SN}, \mathbf{SP}} \ln \mathcal{L}(\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n, \mathbf{GT} | \mathbf{SN}, \mathbf{SP}), \quad (4.4)$$

where $\hat{\mathbf{SN}}$ and $\hat{\mathbf{SP}}$ are the optimised expert annotation performances. Thus, the final ground truth is the binary map \mathbf{GT} that produce the maximum agreement with the expert annotations \mathbf{D}_i in terms of sensitivity and specificity. To solve the simultaneous estimation of the ground truth and the expert annotation performances, the expectation maximisation (EM) algorithm is used. The main advantage of STAPLE over other methods is that it also estimates the ground truth quality.

In comparison to the previous methods, the truth estimation from self distances (TESD) proposed by Biancardi and Reeves [21] provides a divergent approach. The algorithm first computes an Euclidean distance transform for each connected region in the expert annotations \mathbf{D}_i , where the pixels inside the region are assigned with positive distances that increase when moving from the region border to the region centre while outside pixels are assigned negative distances that decrease when moving away from the border. Based on the computed distances, the pixels are weighted and labelled into four categories: inner core, inside border, outside border and outer space. The weighted label maps of each expert are then combined pixel-wise using the centre of gravity method and thresholded to produce the final ground truth map \mathbf{GT} . Note that the distance transform allows to assign confidences also to pixels just outside the annotation border which is not possible in the other described estimation methods.

WEIGHTED EXPERT AREA INTERSECTION

In the case of screening, it is important to detect the diseased test subjects while minimising the falsely detected normal test subjects. In order the image analysis algorithm to detect such test subjects automatically, the ground truth should be representative. In general, the mutual expert agreement is justified criteria to estimate ground truth from multiple expert information, but perhaps not the best for screening purposes. For example, the previously described methods would not include annotations to ground truth if the image had only annotations from single expert out of n experts. If that single expert is, however, truly confident that the annotations are indeed correct, it is possible that other experts just missed those image areas. Therefore, the regions enclosed by the annotations should be considered suspicious in the screening, and the image should be detected and evaluated. Based on this reasoning an alternative ground truth estimation method is presented.

In the DIARETDB1 database, the available expert information is the following (Fig. 4.2): 1) spatial coverage, 2) representative point(s), and 3) confidence level. The spatial coverage consists of circles, ellipses, and polygons which loosely enclose a finding, and a single finding may be represented by multiple primitives. The representative points are distinctive “cue locations” that attracted the expert’s attention. The confidence level describes the expert’s subjective confidence that the annotation represents the given lesion type, where the used confidence scale, $\{< 50\%, > 50\%, 100\%\}$, was collectively agreed

by the medical experts. Furthermore, the subjective expert confidences are converted to numerical values as follows: $\{< 50\%, > 50\%, 100\%\} \rightarrow \{\frac{1}{3}, \frac{2}{3}, 1\}$. For the ground truth estimation, the spatial coverage data and the subjective expert confidences were selected from the available information. As an example, four independent expert annotations for the same lesion type are illustrated in Fig. 4.3.

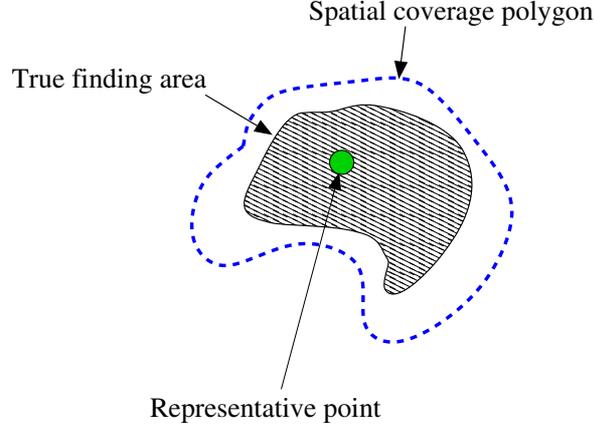


Figure 4.2: The available expert information in the DIARETDB1 database. The expert’s subjective confidence for the annotation is defined as follows: 100%, $> 50\%$ and $< 50\%$.

Consider an eye fundus image for which n ophthalmologists annotated diabetic lesions that should be automatically detected. Furthermore, let the confidence map $\mathbf{D}_{\xi_i}^j$ represent the annotations of j th expert, where the pixel value $D_{\xi_i}^j(x, y)$ denotes the expert’s subjective confidence that the lesion type ξ_i is present at pixel location (x, y) . In comparison to previous, the task is now to find such binary map \mathbf{GT} using the expert annotations $\mathbf{D}_{\xi_i}^j$ that would best describe the object locations that should be detected in the screening of diabetic retinopathy. Since the colleague consultation is the predominant working practise among medical experts, the presented ground truth is also based on the mean expert opinion. However, the averaging is restricted to medical experts that have made annotations in the image, instead of computing the mean opinion over all the experts that participated the annotation. Thus, the mean opinion image can be defined as follows:

$$MOP_{\xi_i}(x, y) = \frac{1}{\hat{n}_{\xi_i}} \sum_{j=1}^{\hat{n}_{\xi_i}} D_{\xi_i}^j(x, y), \quad (4.5)$$

where $MOP_{\xi_i}(x, y)$ is the mean opinion value that indicates the confidence for the lesion type ξ_i to be present in the pixel location (x, y) , and \hat{n}_{ξ_i} is the number of experts that have made annotations in the image for lesion type ξ_i . The final ground truth \mathbf{GT} is produced by thresholding the mean opinion image. The expert annotations were collected using an annotation tool that allows to make overlapping annotations and the overlap is a problem for the presented ground truth estimation if the same expert annotates the

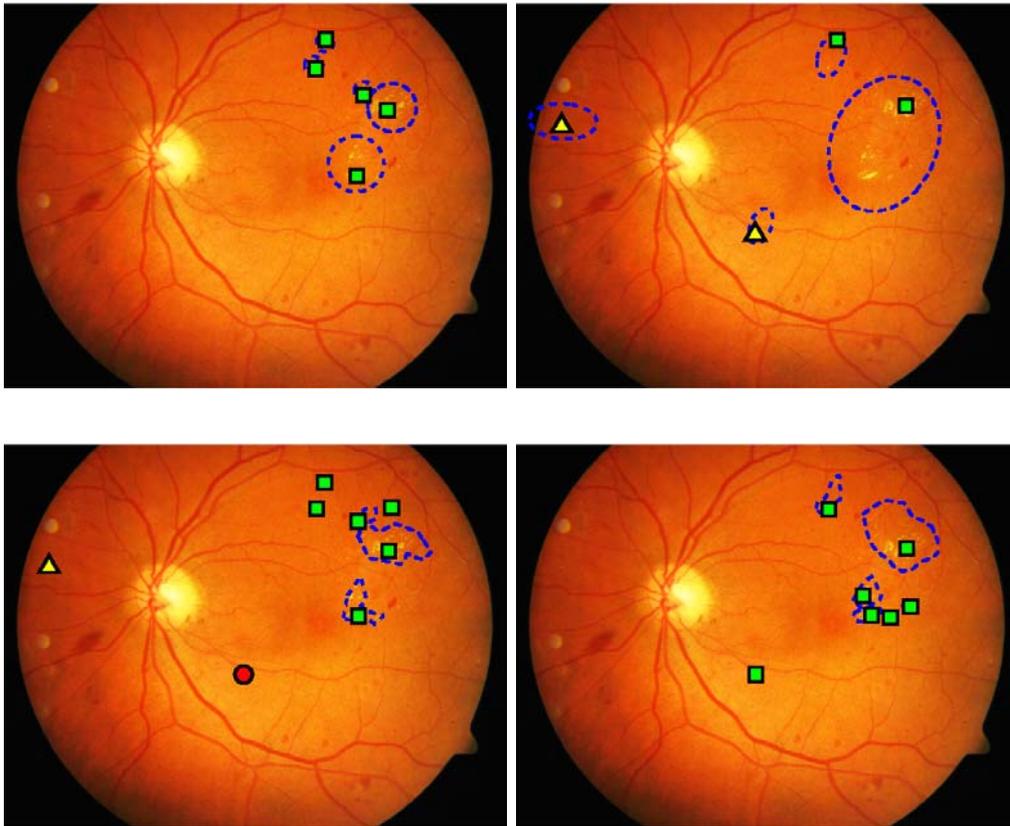


Figure 4.3: Four independent expert annotations (contours and representative points) for the same lesion type (hard exudates). The representative point markers denote the confidence level (*square* = 100%, *triangle* > 50% and *circle* < 50%).

same pixel twice or more by assigning each time the same lesion type label and different confidence. For those pixels, the mean confidence is computed to produce the pixel value $D_{\xi_i}^j(x, y)$. The ground truth estimation using the weighted expert area intersection is illustrated in Fig. 4.4.

The pixel-wise ground truth for the DIARETDB1 database is derived using the fixed mean opinion threshold value 0.75, and if pixels exist in the image after conducting ground truth estimation for certain lesion type then that particular lesion type is considered to be present in the image (image-level ground truth). The selected threshold denotes that if the image comprise annotations only from single expert the subjective confidence for the annotation has to be 100% for it to be in the ground truth. Otherwise, if the image comprise annotations from multiple experts then their mean opinion is required to be at least the given threshold.

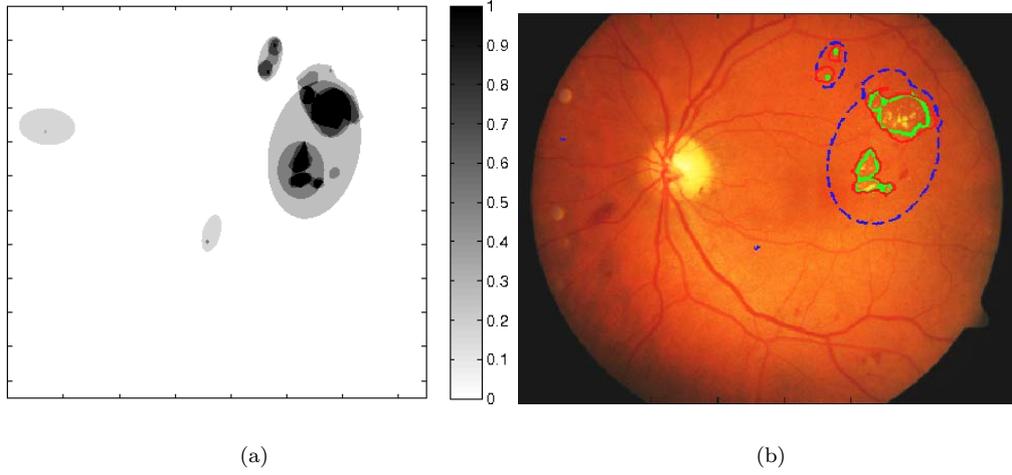


Figure 4.4: Ground truth estimation using the weighted expert area intersection and the annotations shown in Fig. 4.3: a) mean opinion image (darker value denotes higher confidence) b) weighted expert area intersection (blue denotes the mean opinion image threshold 0.25, red for 0.75, and green for 1.00).

4.4 Classifier training from multiple expert information

The multiple expert information also generates the question on how the annotations should be combined for training a classifier. For the expert information available in DIARETDB1 database there exists three intuitive solutions: i) fixed size neighbourhoods of representative points for given confidence levels in Fig. 4.5(a), ii) mean opinion image thresholded by a fixed mean opinion (weighted expert area intersection) in Fig. 4.4(b), and iii) a combination - representative points masked with the result of weighted expert area intersection in Fig. 4.5(b). As reported next, the training strategies combined with all the possible overall scoring strategies are experimentally evaluated to find the best solution for the baseline algorithm.

4.5 Experimental results and discussion

The experiments were conducted on publicly available DIARETDB1 database. As shown in Fig. 4.1, the different classifier training and overall scoring strategies were evaluated utilising the baseline algorithm and the benchmarking framework (Chapter 3). Firstly, the baseline algorithm was trained using the colour information extracted from the image locations produced by the selected classifier training strategy (Fig. 4.7). Secondly, the colour pixels in the test images were classified based on the trained baseline algorithm which resulted a full likelihood map for each tested lesion type (Fig. 4.9). Finally, the overall image scores were computed from the likelihood maps using the proposed overall scoring strategies.

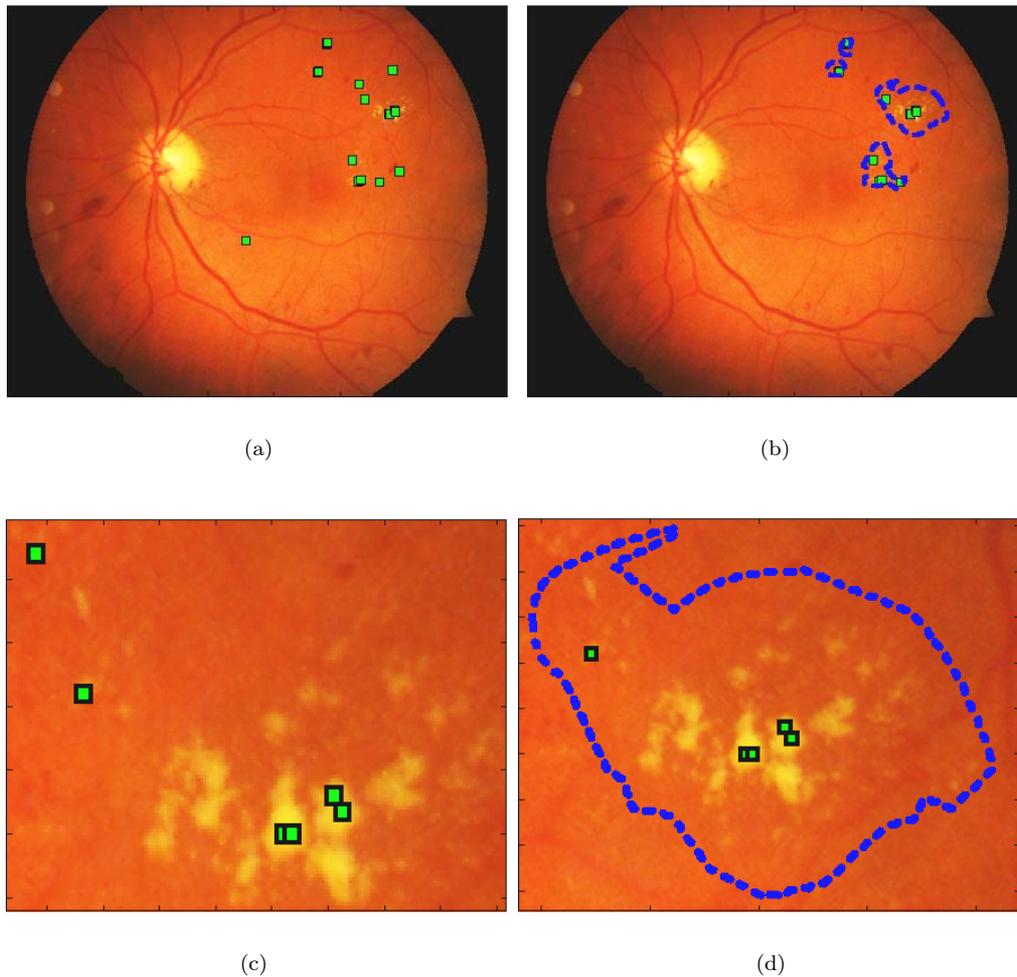


Figure 4.5: Classifier training strategies based on the representative point neighbourhoods (derived for the annotations shown in Fig. 4.3): (a) representative point neighbourhoods (5×5); (b) representative point neighbourhoods masked with the area corresponding 0.75 confidence; (c) close-up of representative points; (d) close-up of masked representative points.

The performance of each setting was evaluated using the computed overall image scores with the image-based evaluation protocol described in the benchmarking framework. The results are qualitatively based on the ROC graphs and quantitatively on the ROC-based quality measure EER (equal error rate) which both are introduced in the benchmarking framework as well. For the image-based evaluation the ground truth was estimated as described in the weighted expert area intersection.

The results in Table 4.1 indicate that better results were always achieved using the

weighted expert area intersection training strategy than using the representative point neighbourhood methods. Apparently, the weighted expert area intersection areas covered better the lesion colour variation required for the classification than representative points which were concentrated only near the most salient lesion points. The results in Table 4.1 also show that the summax rule always produced either the best results or results comparable to the best results. Since the best performance was achieved using the weighted expert area intersection training strategy for which the pure sum (mean), max and product rules were clearly inferior to the summax, the summax rule is preferred. Example ROC curves for the weighted expert area intersection training strategy are shown in Fig. 4.6. The performance of the baseline algorithm with different parameter values of summax rule is shown in Fig. 4.8.

Table 4.1: Equal error rate (EER) for different classifier training and overall scoring strategies.

WEIGHTED EXPERT AREA INTERSECTION								
	0.75				1.00			
	max	mean	summax ($X = 1\%$)	prod	max	mean	summax ($X = 1\%$)	prod
HA	0.25	0.30	0.30	0.35	0.53	0.38	0.40	0.48
MA	0.46	0.43	0.43	0.43	0.39	0.36	0.36	0.43
HE	0.32	0.27	0.27	0.25	0.22	0.25	0.25	0.25
SE	0.27	0.36	0.18	0.36	0.66	0.28	0.30	0.28
Avg.	0.33	0.34	0.29	0.35	0.45	0.32	0.33	0.36
REP. POINT NEIGHBOURHOOD								
	1x1				3x3			
	max	mean	summax ($X = 1\%$)	prod	max	mean	summax ($X = 1\%$)	prod
HA	0.65	0.48	0.48	0.73	0.70	0.43	0.43	0.68
MA	0.71	0.46	0.46	0.46	0.64	0.46	0.46	0.46
HE	0.30	0.20	0.25	0.25	0.15	0.20	0.25	0.30
SE	0.36	0.36	0.36	0.36	0.46	0.36	0.36	0.36
Avg.	0.51	0.38	0.39	0.45	0.49	0.36	0.38	0.45
	5x5				7x7			
	max	mean	summax ($X = 1\%$)	prod	max	mean	summax ($X = 1\%$)	prod
HA	0.60	0.48	0.48	0.68	0.70	0.38	0.53	0.68
MA	0.68	0.43	0.43	0.46	0.46	0.43	0.46	0.43
HE	0.25	0.20	0.20	0.22	0.25	0.25	0.27	0.20
SE	0.38	0.36	0.36	0.55	0.46	0.36	0.28	0.36
Avg.	0.48	0.37	0.37	0.48	0.47	0.36	0.38	0.42
REP. POINT NEIGHBOURHOOD MASKED (AREA 0.75)								
	1x1				3x3			
	max	mean	summax ($X = 1\%$)	prod	max	mean	summax ($X = 1\%$)	prod
HA	0.65	0.48	0.57	0.73	0.65	0.40	0.48	0.68
MA	0.64	0.46	0.50	0.43	0.61	0.50	0.46	0.43
HE	0.40	0.25	0.20	0.20	0.27	0.20	0.20	0.25
SE	0.54	0.28	0.30	0.36	0.22	0.28	0.28	0.36
Avg.	0.56	0.37	0.39	0.43	0.44	0.35	0.36	0.43
	5x5				7x7			
	max	mean	summax ($X = 1\%$)	prod	max	mean	summax ($X = 1\%$)	prod
HA	0.65	0.50	0.43	0.68	0.73	0.48	0.48	0.68
MA	0.52	0.43	0.43	0.43	0.55	0.43	0.50	0.43
HE	0.25	0.27	0.25	0.22	0.25	0.30	0.22	0.25
SE	0.22	0.30	0.28	0.28	0.46	0.28	0.30	0.27
Avg.	0.4100	0.37	0.35	0.40	0.49	0.37	0.37	0.41

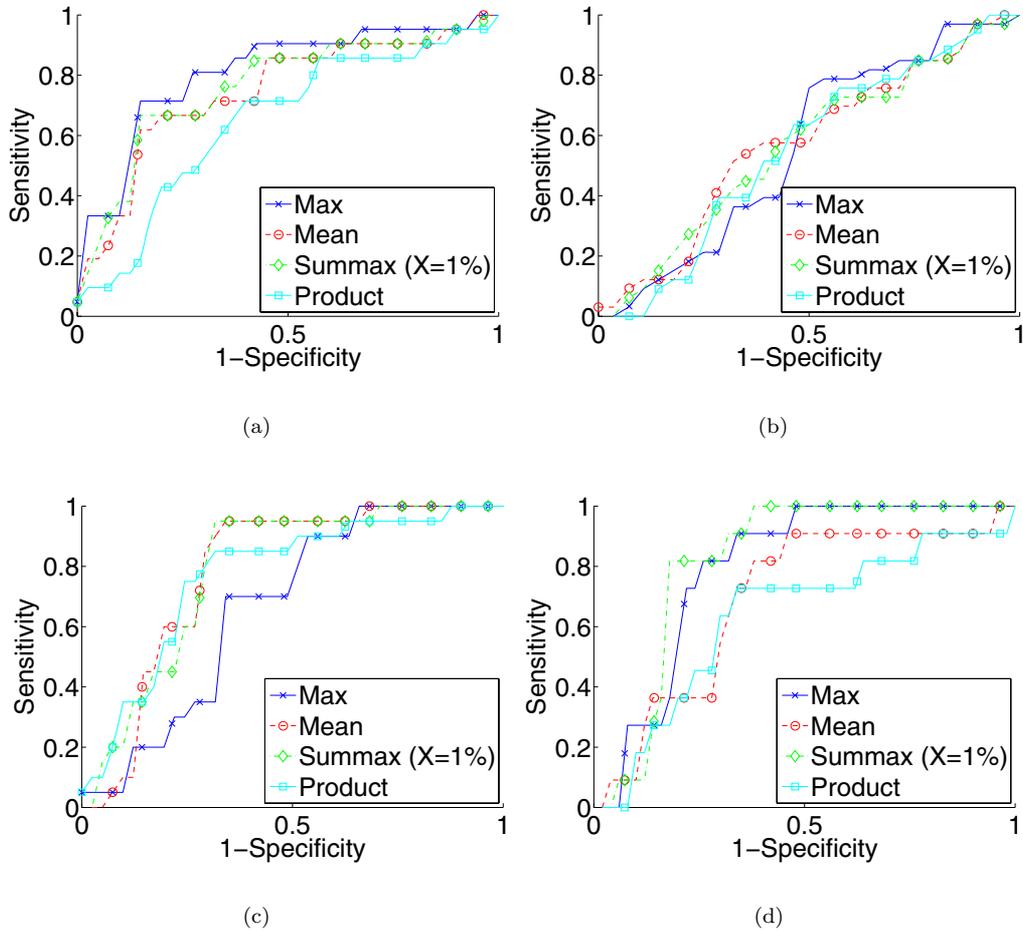


Figure 4.6: Baseline algorithm produced ROC curves for the different overall scoring rules and the weighted expert area intersection training strategy (mean opinion threshold 0.75): (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

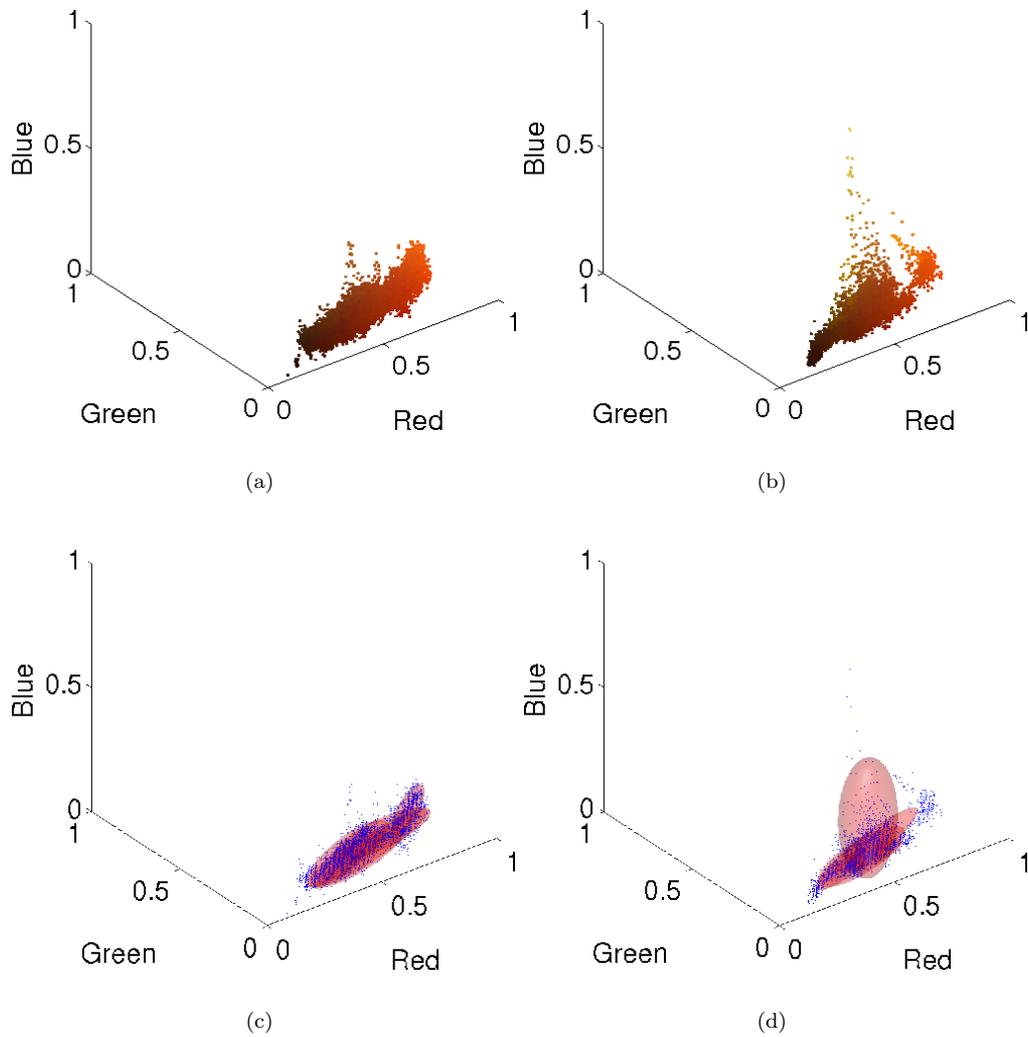


Figure 4.7: Training data produced by the weighted expert area intersection training strategy for haemorrhages (HA) and hard exudates (HE) (points plotted using the true pixel colour): (a-b) colour pixel values of HA and HE in the training set of DIARETDB1; (c-d) estimated colour distributions of HA and HE using the GMM-FJ.

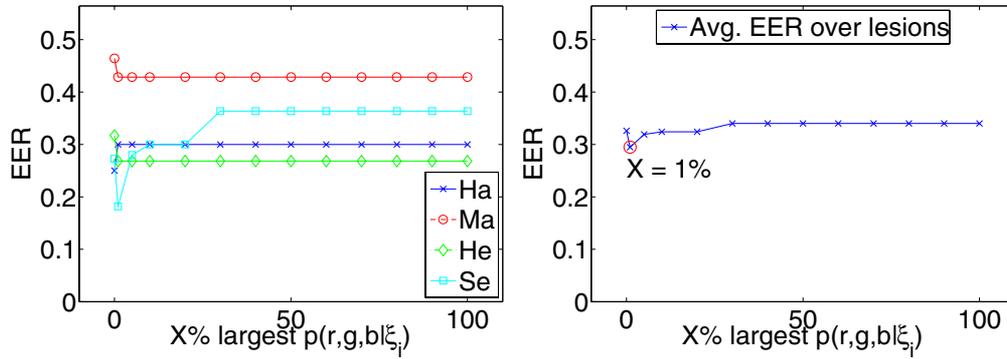


Figure 4.8: Summax overall scoring rule with different parameter values and with the weighted expert area intersection training strategy: (a) baseline algorithm produced equal error rates for each lesion type; (b) baseline algorithm produced equal error rates averaged over the lesion types.



Figure 4.9: Baseline algorithm produced likelihood map using the weighted expert area intersection training strategy: (a) original image (haemorrhages are the red blood areas leaked from the capillary); (b) baseline algorithm produced likelihoods for the haemorrhage areas in the original image.

4.6 Conclusion

In this chapter, two problems essential for supervised learning and classification in eye fundus image processing were addressed: 1) how to combine medical annotations collected from several medical experts into unambiguous ground truth and for training a classifier, and 2) how to form an image-wise overall score for automatic image-based decision making. Firstly, the different alternatives for estimating the ground truth, training a classifier and forming an overall image score were discussed in the context of eye fundus image analysis, and all the training and overall scoring strategies were then experimentally verified against the DIARETDB1 database. The ground truth for the experiments was determined using the proposed weighted expert area intersection. The experimental results showed that the best training strategy for baseline algorithm is the weighted expert area intersection and the best overall scoring strategy is the summax rule ($X = 1\%$). As a conclusion, the full baseline algorithm is shown in Fig. 4.10.

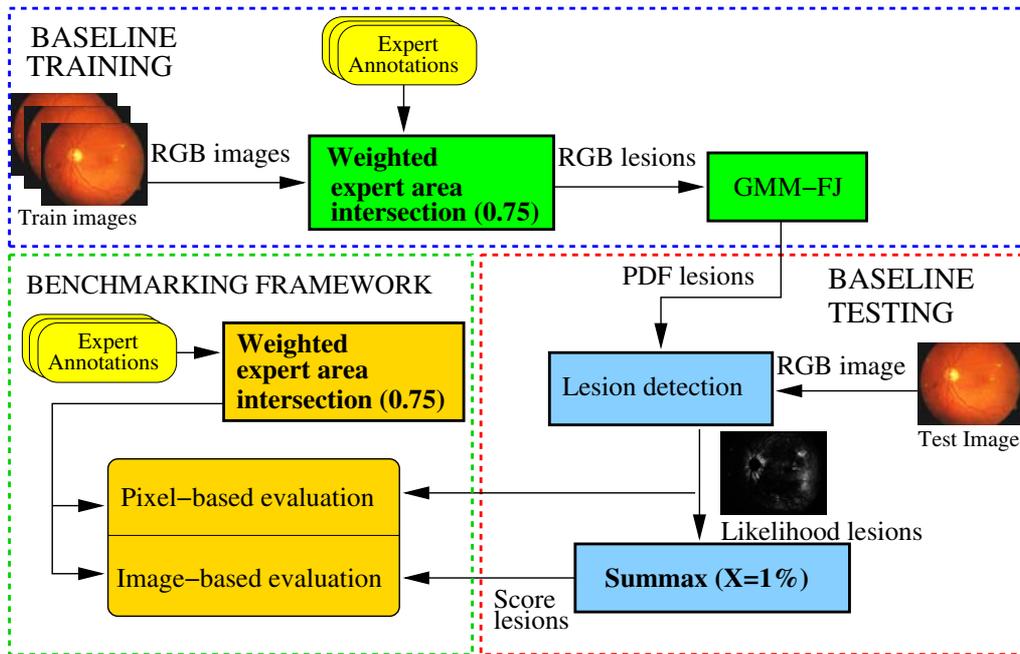


Figure 4.10: Baseline algorithm in full for DIARETDB1 database.

Photometric cue in lesion detection

Eye fundus images are typically inspected by ophthalmologists and their main goal is to detect lesions based on morbid changes in the colour, texture, or shape of tissue or organs. From the available visual cues, the colour or photometric information is one of the simplest features for lesion detection in eye fundus image analysis. The use of colour, however, is often overlooked and the main efforts are put on other cues such as shape and texture. In this chapter, the use of photometric information in detection of diabetic retinopathy is investigated. The purpose is to study how well the lesions can be detected by solely using photometric information, and how the photometric information should be properly utilised in the detection. To investigate these questions, experiments are conducted with the colour space selection, and both illuminance and colour correction. In addition, the use of background class information to support the lesion detection is discussed. The experimental results are based on applying the previously described baseline algorithm and the benchmarking framework on the DIARETDB1 database.

5.1 Introduction

Photometric information, which is defined as the information resulting from the measurement of light, is one of the simplest features for image processing. It affects as N-bit brightness values in its simplest form, trichromatic colour vectors (e.g., RGB) in its most typical form, or spectral reflectance curves in its most complete form (Fig. 5.1). The term “colour” can be interchangeably used with photometric information despite it is typically referring to the reflectance or chromaticity values and omitting the intensity. The photometric information is a primitive feature that is typically considered to require the support of higher level features, such as shape or texture. However, in visual eye fundus image analysis, the colour has a predominant role that could not be replaced by any other cue.

In eye fundus images, the photometric information is the photometric content of an image or images set, where the photometric cue is the photometric evidence of particular interest

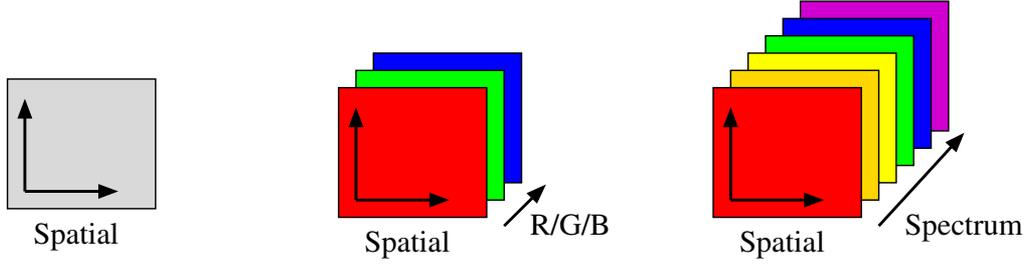


Figure 5.1: Photometric information: (left) monochromatic image (e.g., fluorescein angiogram); (middle) trichromatic colour image (e.g., colour eye fundus image); (right) spectral image.

e.g., colour of the lesions. To study how these photometric cues in the images should be utilised to detect the diabetic lesions, the previously described baseline algorithm (Section 3.6 and Chapter 4) and the benchmarking framework (Chapter 3) are used. In an experimental setting, the photometric cue of the lesions is extracted from the medical expert annotated image locations and using the baseline algorithm the lesions are modelled based on their photometric cue. The trained baseline algorithm is used in classification of the photometric information in the test images that results in for each test image the likelihood maps of the tested lesion types. The pixel value in the likelihood map denotes the probability that certain lesion type is present in that image location. To evaluate the performance and justify the selections made, the image- and pixel-based evaluation approaches of the benchmarking framework are applied to the likelihoods and likelihood derived image scores. The experimental setting is further modified by adding two steps prior to the baseline algorithm: 1) a pre-processing step to investigate the colour space selection and both image illuminance and colour correction and 2) a step to incorporate the background class information (background cue) to support the final decision making. The experimental setting is illustrated in Fig. 5.2.

As the main restriction, the trichromatic values of photometric information are solely considered. The main motivation for this arises from the fact that medical images are almost without exception either monochromatic (gray level) or trichromatic colour images, although spectral information would be the ideal representation and most accurate basis for the analysis. As the second restriction, the experimental part focuses on eye fundus images and detecting lesions indicative of diabetic retinopathy. Moreover, the selections in the experimental part are based on the quantitative image-based evaluation results since the image-based evaluation approach, i.e. lesion type is present or not present in the image, is considered to correspond the “subject-wise” decision making in the clinical practice. The pixel-based evaluation results are also reported as they are highly informative for the method development and an important addition to the image-based results. The experiments were conducted on two image databases: 1) the DIARETDB1 database to conduct the main analysis, and 2) the BRISTOLDB database to study the applicability of the DIARETDB1 database results.

The image illuminance correction described in this chapter is reported in [90] and the

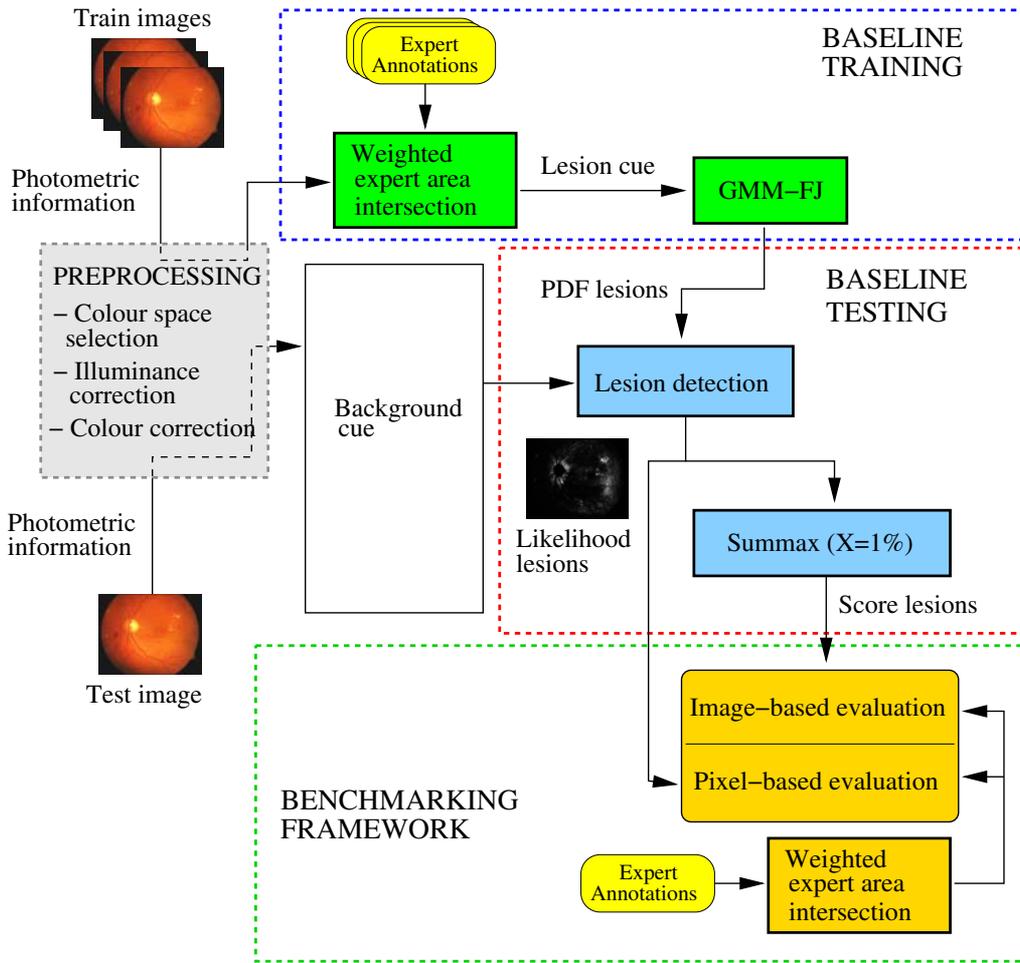


Figure 5.2: Experimental setting to investigate the use of photometric cues in lesion detection.

chapter is organised as follows. Section 5.2 briefly reviews the studies that have investigated the use of photometric information in detection of diabetic retinopathy. Section 5.3 discusses the trichromatic photometric information in eye fundus images. Section 5.4 covers the colour space selection. Section 5.5 studies the use of background class information. Section 5.6 discusses the non-uniform image illuminance and its correction. Section 5.6 investigates the use of colour correction. Section 5.8 discusses the applicability of the results. Section 5.9 studies the influence of under and overexposed pixels in the experiments. Finally, the conclusions are drawn in Section 5.10.

5.2 Related work

During the last decade, a considerable amount of effort has been spent on the automated detection of diabetic retinopathy from digital eye fundus images. As a result, a number of image processing and computer vision techniques have been published in the literature. The relevant research was given in Section 2.5 which encapsulated the main algorithms used in the field during the past 10-15 years. Despite the numerous elaborate algorithms; the strength of colour cue is not very often studied, although it is regarded as a powerful descriptor with substantial potential [188, 127]. One of the few studies that examine the strength of colour cue for diagnosis of eye fundus images was conducted by Truitt et al. in [168], where RGB colour information was studied by analysing the class distribution properties (histograms) of anatomically normal and pathological areas in digitised retinal images. It was concluded that despite an overlap exist in colour of retinal tissues, it may be possible to separate diabetic lesions from the non-lesion areas solely based on colour information.

5.3 Trichromatic photometric information

In colour eye fundus photography, the image of retina is created by emitting a flash lamp produced light into patient's eye through complex fundus camera optics and capturing the reflected light to the camera sensor. If the emitted light is characterised with the power spectrum $L(\lambda)$, the reflective properties of the retina with the reflectance function $S(\lambda)$ then the reflected light $C(\lambda) = L(\lambda) \cdot S(\lambda)$ enters the camera, where it is sampled by the camera sensor. In the RGB camera, the photometric information is reproduced by combining three colour channels (trichromacy) in which the RGB values are given by [103],

$$R = \int R(\lambda)L(\lambda)S(\lambda)d\lambda, \quad (5.1)$$

$$G = \int G(\lambda)L(\lambda)S(\lambda)d\lambda, \quad (5.2)$$

$$B = \int B(\lambda)L(\lambda)S(\lambda)d\lambda, \quad (5.3)$$

where $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$ define the spectral sensitivity of the camera sensor and the integral covers (λ) the visible spectrum ranging from 400 to 700 nanometres. The light in the camera is divided into RGB colour components either using colour filter arrays (e.g. Bayer filter) or beam splitters (e.g. three-charge-coupled devices, 3CCD). In the first option, a filter mosaic is set over the photosensitive elements of the camera sensor to capture the colour information. In the second option, the incoming light is divided and directed into three separate camera sensors using a prism assembly.

In the RGB representation, RGB values are dependent on the illuminant, imaged target and the camera sensor sensitivities. If the same retina was imaged with two different uncalibrated fundus cameras, the RGB content of the lesions would be different in the images due to changed illumination and camera sensitivity characteristics. In addition, the RGB representation combines the colour (chrominance) and intensity (luminance)

information for which the chromaticity coordinates (normalised RGB) can be derived as follows [103]

$$\hat{r} = \frac{R}{(R + G + B)}, \quad \hat{g} = \frac{G}{(R + G + B)}, \quad \hat{b} = \frac{B}{(R + G + B)}, \quad (5.4)$$

where the image intensity is $R + G + B$ and $\hat{r} + \hat{g} + \hat{b} = 1$.

5.4 Colour space selection

The first concern with trichromatic photometric information is the colour space selection. There are numerous colour spaces available, but the appropriate colour space is typically selected for each application separately since none of the colour spaces is universally better than the others. However, it is assumed that any particular space constructed using a linear transformation from another should not affect the results. Any linear colour space from the RGB values is given by

$$\begin{bmatrix} \overline{R} \\ \overline{G} \\ \overline{B} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (5.5)$$

where the 3×3 matrix defines a linear transformation to the colour space \overline{RGB} .

The search for suitable colour space is further restricted to two issues that can affect the analysis: expert annotations and lesion separability in the colour space. Since the expert annotations have been performed manually by medical doctors using RGB displays, it is motivating to study whether the colour representation has any effect on the analysis. If it had, colour spaces designed to model the human colour vision (such as CIE $L^*a^*b^*$) would represent the ground truth better. Another important issue that directly contributes to the ability of the baseline algorithm to discriminate the lesions is how photometric cues of lesions are distributed in the colour space. To study the addressed issues, two delicately selected colour spaces are used in the experiments: CIE $L^*a^*b^*$ and HSV. Instead of experimenting with all the possible alternatives, the selected colour spaces are used to indicate if further experiments are required.

5.4.1 CIE $L^*a^*b^*$

CIE $L^*a^*b^*$ is a perceptually uniform colour space recommended by the Commission International d'Eclairage (CIE) [27], where L^* denotes the lightness axis and the colour opponents a^* and b^* define the red-green and yellow-blue axes. In theory, a unit Euclidian distance in CIE $L^*a^*b^*$ is the threshold for a human to distinguish colour difference. The colour transformation from XYZ to CIE $L^*a^*b^*$ colour space is a non-linear mapping

and defined as

$$\begin{aligned}
L^* &= 116 f\left(\frac{Y}{Y_N}\right) - 16, \\
a^* &= 500 \left[f\left(\frac{X}{X_N}\right) - f\left(\frac{Y}{Y_N}\right) \right], \\
b^* &= 200 \left[f\left(\frac{Y}{Y_N}\right) - f\left(\frac{Z}{Z_N}\right) \right],
\end{aligned} \tag{5.6}$$

where

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & \text{if } t > 0.008856 \\ 7.787t + \frac{16}{116}, & \text{otherwise} \end{cases}$$

Here X , Y and Z are the red, green and blue for standard observer (CIE 1931 XYZ), and X_N , Y_N and Z_N define the illuminant dependent reference white point in the XYZ colour space. In general, the characteristics of the illuminant should be known when transforming RGB to XYZ values, and further to the $L^*a^*b^*$ colour space. However, if image capturing is conducted using the same camera under the same illuminant then the white point in the transformation can be fixed, e.g. CIE standard illuminant D65.

5.4.2 HSV

HSV is an alternative colour representation for RGB that is based on the perceptual variables: hue (H), saturation (S) and value (V) [154]. The hue and saturation are presented in polar coordinates and the value defines the brightness axis. Although, the colour axes are defined as perceptual variables the HSV colour space does not model human colour vision. The transformation from RGB to HSV colour space is nonlinear mapping and defined as

$$\begin{aligned}
H &= \begin{cases} 0, & \text{if } \Delta = 0 \\ (360 + 60 \cdot \frac{G-B}{\Delta}) \bmod 360, & \text{if } \max(R, G, B) = R \\ 120 + 60 \cdot \frac{B-R}{\Delta}, & \text{if } \max(R, G, B) = G \\ 240 + 60 \cdot \frac{R-G}{\Delta}, & \text{if } \max(R, G, B) = B \end{cases} \\
S &= \begin{cases} 0, & \text{if } V = 0 \\ \frac{\Delta}{V}, & \text{otherwise} \end{cases} \\
V &= \max(R, G, B),
\end{aligned} \tag{5.7}$$

where $\Delta = V - \min(R, G, B)$, $H \in [0, 360[$ and $S \in [0, 1]$.

5.4.3 Experimental results and discussion

The quantitative DIARETDB1 results for the RGB, CIE L*a*b* and HSV colour spaces are shown in the Tables 5.1 and 5.2. The mean ROC curves are plotted in Figs. B.1 and C.1 in Appendices B and C. From the quantitative results in Tables 5.1 and illustrations in Fig. B.1, it is evident that the selection of the colour space has only little or no effect when comparing colour spaces on image-basis. However, in the pixel-based comparison both colour spaces produced better results than RGB colour space (Table 5.2). Since it was specified that the selections in the experiments are based on the quantitative image-based results (Sec. 5.1), it is motivated to continue using the RGB colour space. Any bias related to the colour space does not affect the image-based decision making, or it is compensated by the GMM estimation. This result is well in accordance with the results from other fields (e.g., face detection by skin colour [105]), and supports the finding that Gaussian mixture models are most accurate for data in unnormalised colour spaces. The training error for the baseline algorithm using RGB colour space is shown in Tables D.1 and D.2 in Appendix D, and in Figs. D.1 and D.2 .

Table 5.1: Mean EER results (image-based evaluation) for the baseline algorithm using different colour spaces. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry was conducted three times.

	HA	MA	HE	SE	Avg.
Baseline (RGB)	0.29	0.48	0.24	0.25	0.31
CIE L*a*b*	0.30	0.46	0.25	0.28	0.32
HSV	0.35	0.43	0.25	0.28	0.33

Table 5.2: Mean EER results (pixel-based evaluation) for the baseline algorithm using different colour spaces. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

	HA	MA	HE	SE	Avg.
Baseline (RGB)	0.37	0.36	0.39	0.41	0.38
CIE L*a*b*	0.37	0.36	0.34	0.39	0.36
HSV	0.34	0.35	0.30	0.37	0.34

5.5 Utilising background class information

In the previous experiments, the training was performed only using the colour information of the lesion types. The decision whether a lesion type is present or not in the test image or test image pixel was solely based on the colour information of that specific lesion type while the information concerning other retinal structures and other lesion types was neglected. Thus, no assumptions were made about other retinal structures or lesion types. In this section, both “lesion type present” and “lesion type absent” classes are

modelled instead of just modelling the “lesion type present” class. The objective is to study if the use of background class can support the detection of diabetic retinopathy.

As previously, the detection problem is to determine to which category a given image or pixel belongs, where the possible classes are “lesion type present” and “lesion type absent” (background). In this case, however, both classes have a representative density functions, and therefore the decision for given pixel value can be written as follows:

$$p(\xi_i)p(\mathbf{x}|\xi_i) > p(\bar{\xi}_i)p(\mathbf{x}|\bar{\xi}_i), \quad (5.8)$$

where $p(\xi_i)$, $p(\bar{\xi}_i)$ are the prior probabilities for the lesion type ξ_i and the background class $\bar{\xi}_i$, and $p(\mathbf{x}|\xi_i)$ and $p(\mathbf{x}|\bar{\xi}_i)$ are the corresponding probability density functions with respect to the colour pixel value \mathbf{x} . Equivalently, Eq. 5.8 can be written in form:

$$l_{\xi_i}(\mathbf{x}) = \frac{p(\mathbf{x}|\xi_i)}{p(\mathbf{x}|\bar{\xi}_i)} > \frac{p(\bar{\xi}_i)}{p(\xi_i)}, \quad (5.9)$$

where the term $l_{\xi_i}(\mathbf{x})$ is known as the likelihood ratio. Now, the pixel and image scores can be derived from the likelihood ratios as follows:

$$\begin{aligned} \zeta_{\xi_i}^{pix} &= l_{\xi_i}(\mathbf{x}), \\ \zeta_{\xi_i}^{im} &= \sum_{j \in N} l_{\xi_i}(\mathbf{x}_j), \quad (\text{Summax rule}) \end{aligned} \quad (5.10)$$

where N is the set of $X\%$ of the largest $l_{\xi_i}(\mathbf{x}_j)$. To estimate the density function $p(\mathbf{x}|\bar{\xi}_i)$ that represents the background class, the training data was extracted from randomly selected image locations for which none of the experts labelled to belong to lesion type ξ_i . The training data was sampled instead of using the whole set to conduct GMM-FJ estimation in reasonable time. The used pixel counts per training image were: 100, 1000, 10000. It is also noteworthy to mention that the evaluation in this case searches automatically for each lesion type the ratio of prior probabilities, $P(\bar{\xi}_i)$ and $P(\xi_i)$, that best discriminates the lesioned pixels and images from non-lesioned ones with respect to sensitivity and specificity.

5.5.1 Experimental results and discussion

The quantitative DIARETDB1 results for the use of background class information to support lesion detection are shown in Tables 5.3 and 5.4. The mean ROC curves are plotted in Figs. B.2 and C.2 in Appendices B and C. The quantitative results indicate that the use of background class information indeed improves the pixel-level performance especially for hard and soft exudates, and regardless of the used pixel count. On the other hand, the division in the likelihood ratio easily produced outlier pixels when the likelihoods were small, and since the likelihood behaviour in the image is relatively irregular the outlier pixels were very likely. This clearly affects the image-level performance when the summax rule with 1% of the highest likelihood ratios ($X=1\%$) were used to indicate the presence of certain lesion type in the image. Fortunately, the experiments also showed that the outlier effect can be compensated by increasing the pixel count in density estimation and the proportion of the largest likelihood ratios in the summax rule (e.g.

X=80%). In conclusion, the use of background class information is recommended, since it produced better performance on pixel-level and can achieve comparable performance on image-level.

Table 5.3: Mean EER results (image-based evaluation) for the baseline algorithm using the background class information. The number denote the pixel count (100,1000,10000) per image used in the training of background classes. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

		HA	MA	He	SE	Avg.
	Baseline (RGB)	0.29	0.48	0.24	0.25	0.31
Summax ($X = 1\%$)	Lr100	0.25	0.51	0.37	0.35	0.37
	Lr1000	0.29	0.49	0.35	0.35	0.37
	Lr10000	0.32	0.50	0.25	0.36	0.36
Summax ($X = 80\%$)	Lr100	0.27	0.49	0.36	0.29	0.35
	Lr1000	0.29	0.51	0.32	0.26	0.34
	Lr10000	0.31	0.49	0.24	0.28	0.33

Table 5.4: Mean EER results (pixel-based evaluation) for the baseline algorithm using the background class information. The number denote the pixel count (100,1000,10000) per image used in the training of background classes. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

	HA	MA	HE	SE	Avg.
Baseline (RGB)	0.37	0.36	0.39	0.41	0.38
Lr100	0.36	0.36	0.27	0.37	0.34
Lr1000	0.37	0.36	0.32	0.37	0.35
Lr10000	0.35	0.35	0.27	0.37	0.34

5.6 Image illuminance correction

The imaging process is itself an error source. Specifically, non-uniform image illuminance due to non-uniform illumination on the imaged object and imaging optics is common in eye fundus images. This deficiency is clearly visible in all eye fundus images, where light enters through a relatively small aperture (pupil) to the centre of the eye fundus and part of the intensity is lost towards the fundus borders (Fig. 5.3(a)). A few simple methods for illuminance correction, such as using components a^* and b^* of the CIE $L^*a^*b^*$ colour space (Eq. 5.6) or normalised RGB (Eq. 5.4), do exist. In general, however, proper illuminance correction is always problem-specific and should be based on measurements supporting the general knowledge of optics and physics. The bright centre and dim borders in fundus images are due to the imaged object (the eye fundus with a 3-D concave shape), wide-angle optics of the eye fundus cameras, and geometry of the sensor array. To overcome the non-ideal nature of imaging systems without accurately modelling a specific system, a set of existing correction methods is compared. Most of them are based on known optical imaging system characteristics, such as the \cos^4 law and Kang-Weiss illuminance model, or on indirectly derived information such as the radial, bivariate, elliptic, and polynomial parametric models [90]. Before describing the parametric models,

the relation between the imaging and the image illuminance, and how it is corrected is described.

5.6.1 Imaging and illuminance

The imaging process is modelled as $I(x, y) = f_c(r(x, y) + \epsilon_1) + \epsilon_2$, where $I(x, y)$ is the pixel value of the observed image, $f_c(\cdot)$ is the radiometric response function of the imaging system, $r(x, y)$ is the light received by the pixel which is corrupted by ϵ_1 (shot and thermal noise) and ϵ_2 (quantisation error, amplifier noise, D/A and A/D noise) [169]. To include the non-uniform image illuminance in the imaging model, the term $f_c(r(x, y) + \epsilon_1)$ can be written as $f_c(r(x, y) \cdot \varphi(x, y) + \epsilon_1)$, where $\varphi(x, y)$ is a distortion factor that represent illuminance degradation and $r(x, y) \cdot \varphi(x, y)$ represents pixel-wise multiplication. By assuming small noise level compared to the signal, the terms ϵ_1 and ϵ_2 can be ignored, and the the imaging model simplifies to form

$$I(x, y) = f_c(r(x, y)) \cdot \varphi(x, y) = I_0(x, y) \cdot \varphi(x, y), \quad (5.11)$$

where $I_0(x, y)$ is the distortion free image and $\varphi(x, y)$ is the distortion factor estimated using the model-based correction methods. Note that the imaging model is possible to derive also as an additive model $I_0(x, y) + \varphi(x, y)$ [166]. Empirically, the additive model seems to be less sensitive to estimation errors, whereas the multiplicative model preserves the contrast. In this section, multiplicative model is used.

5.6.2 Illuminance estimation and its correction

The image illuminance is estimated by fitting a parametric model to the image points that represent the image illuminance. The method proposed in [48] assumes that areas which do not contain sudden changes reveal such behaviour. By following this method, background image points that exclude disease, blood vessels and optic disc are extracted from the intensity image by using the standard deviation and average pixel value in a sliding window. If the distance between the centre pixel value and the average pixel value of the sliding window is less than standard deviation, the centre pixel is considered as background. The parameters are recovered by minimising the sum squared distance between the parametric model and the background points. Now, the illuminance corrected image from for Eq. 5.11 can be recovered from the fitted result as follows:

$$I_0(x, y) = \frac{I(x, y)}{\varphi(x, y)} \approx \frac{I(x, y)}{\phi(x, y)}, \quad (5.12)$$

where $\phi(x, y)$ is the fitted parametric model value at image location (x, y) . In the RGB colour space, the natural choice for the background point extraction and the image illuminance estimation is the intensity $[(R + G + B)/3]$, but in the case of eye fundus images the red channel can be used directly [138]. The distortion free RGB image can be recovered simply by dividing pixel-wise each colour channel of the image separately with the same correction factor as follows

$$R_0(x, y) = \frac{R(x, y)}{\phi(x, y)}, \quad G_0(x, y) = \frac{G(x, y)}{\phi(x, y)}, \quad B_0(x, y) = \frac{B(x, y)}{\phi(x, y)}, \quad (5.13)$$

where $R_0(x, y)$, $G_0(x, y)$ and $B_0(x, y)$ are the colour channels of the illuminance corrected image, and $R(x, y)$, $G(x, y)$ and $B(x, y)$ the corresponding channels of the illuminance degraded image. The image illuminance correction is illustrated in Fig. 5.3.

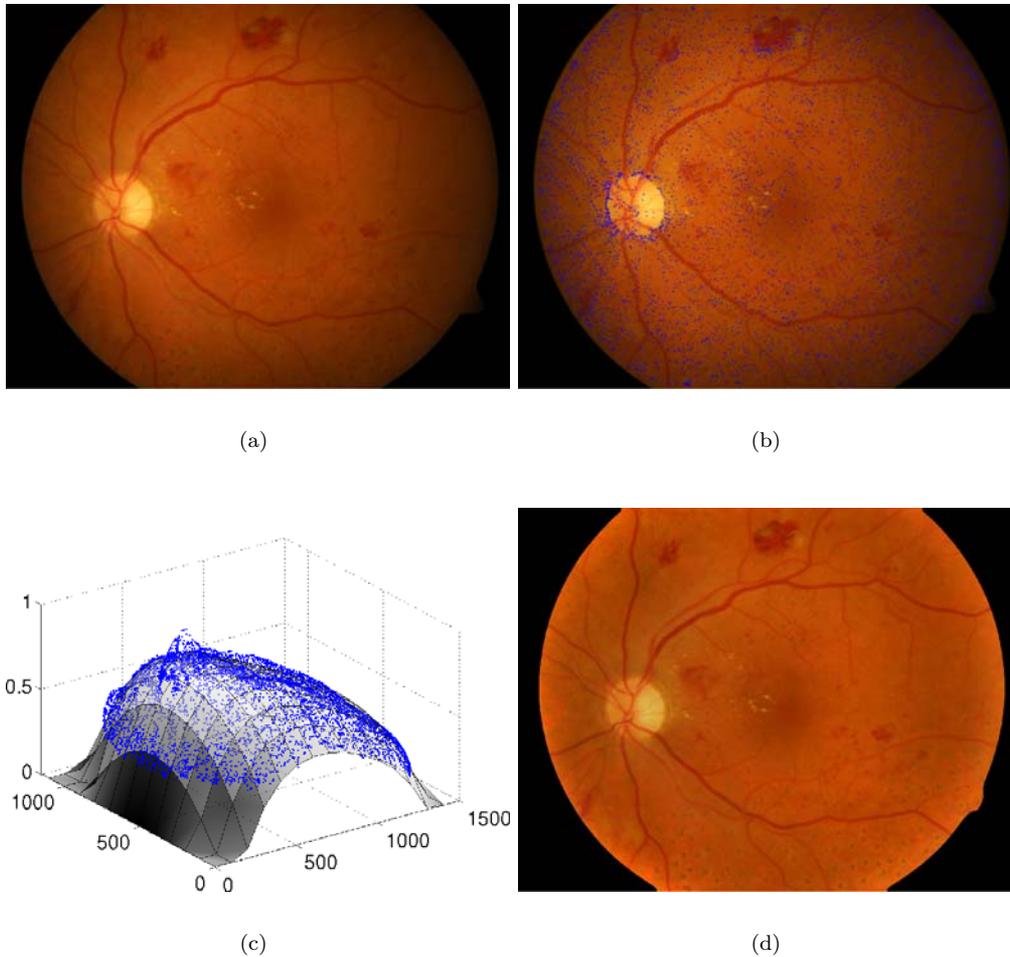


Figure 5.3: Example illuminance correction of an eye fundus image: (a) original image with degraded illuminance; (b) data representing the degraded illuminance (denoted blue in the image) (c) estimating degraded illuminance with a parametric model (bivariate polynomial); (d) illuminance corrected image.

5.6.3 Parametric illuminance models

For simple lenses, the natural vignetting is approximated using the cosine law of illumination [135]. For the pixel (x, y) with distance r from the image centre, the cosine law can be formulated in a form as

$$\phi_c(x, y) = L_0 \cos^4(\tan^{-1}(r/f)), \quad (5.14)$$

where f is the effective focal length, L_0 is the (maximum) irradiance at the image centre. The default power for the cosine is four as in Eq. 5.14, but corrective measures applied to the wide-angle optics reduce the power.

A more elaborate model based on physical consideration was proposed by Kang and Weiss [84]. In addition to radial falloff G_k , the model includes the off-axis illumination A_k and the observed object tilt T_k [84]. For the pixel (u, v) with the distance r from the image centre $(u, v) = (0, 0)$, the model is formulated as $\phi_k(u, v) = A_k G_k T_k$, where

$$\begin{aligned} A_k &= \frac{1}{(1 + (r/f)^2)^2}, \\ G_k &= (1 - \alpha_0 r), \\ T_k &= \cos \tau \left(1 + \frac{\tan \tau}{f} (u \sin \chi - v \cos \chi) \right)^3. \end{aligned} \quad (5.15)$$

f is the effective focal length, α_0 represents the radial vignetting factor coefficient, τ describes the rotation of the observed planar object around the axis parallel to the optical axis, and χ describes the rotation of the observed planar object around the x -axis. To compare the radial polynomial model and the Kang-Weiss model, the radial falloff term G_k is replaced with $\phi_r(x, y)$ from Eq. 5.16.

The radial polynomial model is a common model for approximating symmetrical radial falloff of intensity with increasing distance from the image centre (e.g., in image mosaicing [91]). The radial falloff for the pixel (x, y) is defined as

$$\phi_r(x, y) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \dots + \alpha_n r^n, \quad (5.16)$$

where r is the distance from the image centre and $\alpha_0, \alpha_1, \dots, \alpha_n$ are the estimated parameters.

A more general approach to estimate the intensity falloff, also applied for eye fundus images in [109], is to use polynomial functions which are not necessarily concentric with the image centre and do not require symmetry [138]. A polynomial up to the second order for the pixel (x, y) is defined as

$$\phi_p(x, y) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy, \quad (5.17)$$

where $\beta_0, \beta_1, \dots, \beta_5$ are the estimated parameters.

The empirical model proposed for retinal illumination correction in [64] approximates vignetting and uneven illumination with elliptic paraboloids which are allowed to shift, scale and rotate. The elliptic paraboloid for the pixel (x, y) is defined as

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ \phi_e(x, y) \end{bmatrix} = R_y(p_9) R_x(p_8) R_z(p_7) \begin{bmatrix} x - p_1 \\ y - p_2 \\ z \end{bmatrix}, \quad (5.18)$$

where

$$z = p_6 \left(\frac{(x - p_1)^2}{p_4} + \frac{(y - p_2)^2}{p_5} \right) + p_3 . \quad (5.19)$$

R_y, R_x, R_z correspond to the rotation matrices around the corresponding coordinate axis and $p_1, p_2 \dots p_9$ are the estimated parameters. In [64], both the intensity falloff and luminosity glare in the periphery of the image were modelled. However, in this study only the intensity falloff towards the image border is considered. The illuminance estimation of image in Fig. 5.3 using the parametric models are illustrated in Fig. 5.4.

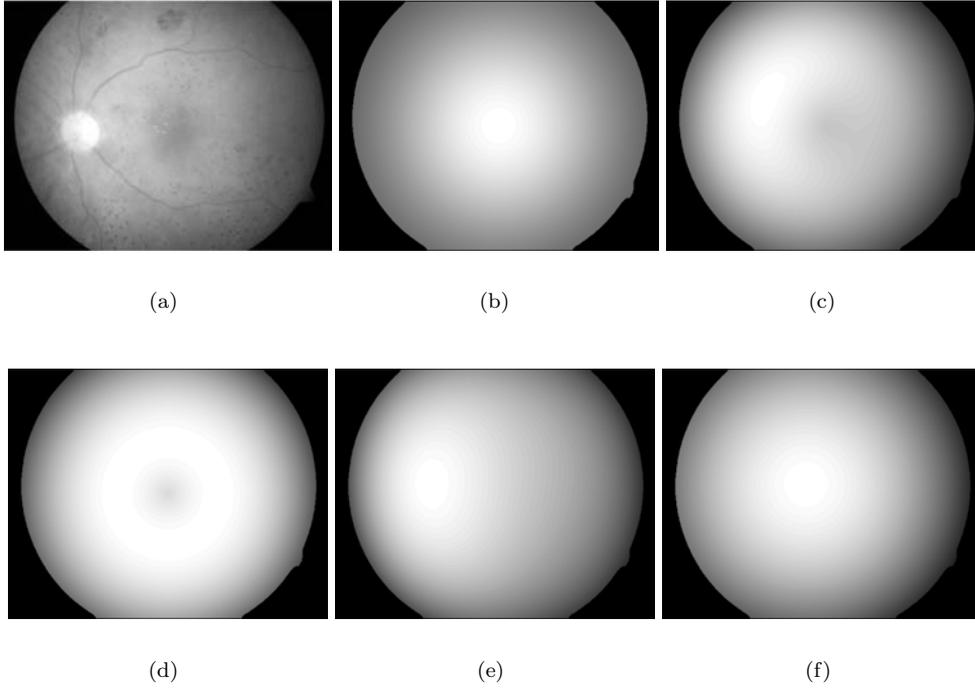


Figure 5.4: Examples of estimated image illuminances using the parametric models: (a) Image used in the illuminance estimation (red channel) (b) \cos^4 law of illumination; (c) Kang-Weiss model; (d) 3rd order radial polynomial; (e) 4th order bivariate polynomial; (f) elliptic paraboloid.

5.6.4 Experimental results and discussion

The quantitative DIARETDB1 results for the illuminance correction are shown in the Tables 5.5 and 5.6. The mean ROC curves are visualised in Figs. B.3 and C.3 in Appendices B and C. The quantitative results given in the Table 5.5 show that the simple heuristic methods, ab (from Lab) and rg (from normalised RGB), decreased the performance, and most other methods had a negligible effect (\cos^4 , elliptic, Kang-Weiss, radial). Although, the results were not promising, the results in Fig. B.3 states that

the best method (fourth order polynomial) improves the detection in the high specificity range, which is important from the clinical point of view, and otherwise the performance is comparable to the baseline algorithm or better. The results indicate that illuminance correction could indeed be helpful if correctly applied, where the improvement would be accomplished in the spatial domain rather than at image level (Fig. F.1 in Appendix F). This can be explained by the fact that the correction reduces the colour variance, but the relative difference between each lesion and background is not substantially affected. The results also address that image illuminance correction is task specific and the general approaches presented in this thesis are not efficient in the case of eye fundus images.

In general in the eye fundus imaging, the illuminance correction does not substantially affect the detection accuracy in good imaging conditions since machine learning methods such as the GMM-FJ estimation can compensate the elementary non-idealities. It is clear that the fourth order polynomial slightly improves the results, but due to its general nature it is not optimal for the task-dependent correction procedure. To keep the baseline as general as possible it is preferred not to include the illumination correction; both the standard heuristic methods and the optics-based methods (\cos^4 , order 3 radial, elliptic and Kang-Weiss) failed to improve the results.

Table 5.5: Mean EER results (image-based evaluation) for the baseline algorithm using illuminance correction. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

	HA	MA	He	SE	Avg.
Baseline (RGB)	0.29	0.48	0.24	0.25	0.31
a*b* (CIE L*a*b*)	0.29	0.55	0.32	0.28	0.36
rg (R+G+B=1)	0.29	0.39	0.30	0.36	0.34
\cos^4	0.36	0.45	0.25	0.28	0.34
Radial poly (ord. 3)	0.33	0.46	0.22	0.28	0.32
Elliptic parab.	0.30	0.42	0.25	0.31	0.32
Kang-Weiss (ord. 3)	0.33	0.36	0.25	0.31	0.31
Polynomial (ord. 4)	0.33	0.35	0.25	0.28	0.30

Table 5.6: Mean EER results (pixel-based evaluation) for the baseline algorithm using illuminance correction. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

	HA	MA	He	SE	Avg.
Baseline (RGB)	0.37	0.36	0.39	0.41	0.38
a*b* (CIE L*a*b*)	0.39	0.38	0.36	0.40	0.38
rg (R+G+B=1)	0.37	0.38	0.32	0.39	0.37
\cos^4	0.39	0.37	0.32	0.43	0.38
Radial poly. (ord. 3)	0.38	0.39	0.32	0.42	0.38
Elliptic parab.	0.36	0.36	0.31	0.37	0.35
Kang-Weiss (ord. 3)	0.36	0.36	0.31	0.38	0.35
Polynomial (ord. 4)	0.35	0.36	0.33	0.40	0.36

5.7 Colour correction

The photometric information should be consistent over all the images representing the same finding or landmark, but there are several natural sources of photometric variation in eye fundus images: changing of the imaging device or its parameters (e.g., eye fundus camera), changing of the illumination source (spectral changes depending on the illumination source and time), and normal variation between the individuals (including iris, lesions and tissues affecting the imaging). From the computational and clinical viewpoint, the classifier should cope with the normal variation, but other inconsistencies should be minimised. The best solution is to calibrate the imaging system. Unfortunately, eye fundus images are typically captured using uncalibrated imaging systems with unknown imaging parameters, and one cannot rely on the photometric consistency between the images. Therefore, known techniques for normalising the colours are applied including an additional novel approach.

The colour correction is focused on reference-based methods, where the images are mapped to a common colour space that is expected to minimise the colour variance within a lesion type. Therefore, the aim is not only to ensure that the colour remains the same under varying illumination, but also to control the normal colour variations such as pigmentation. A set of well-known colour normalisation techniques such as histogram matching, and comprehensive colour normalisation [42] were compared in [58] for reducing the photometric variance between retinal colour images. Since the histogram specification was considered the best method, it was implemented [67] and included in the study along with a similar reference-based method known as colour transfer [134]. In addition, a straightforward method is proposed that is based on retinal landmark matching using a homography which estimates the linear transform between a template and an input image with annotated landmarks. The typical landmarks for the purpose are optic disc, fovea, blood vessels, fundus, and the black borders surrounding the image. Using the transform, all the pixel values can be transformed into the template colour space. As a comparison, the first two methods derive the reference colour space from colour content of a single image, whereas the last-mentioned constructs a mean reference space based on the colour information extracted from multiple images. Next, the methods are described in detail.

5.7.1 Histogram matching

Histogram matching [67], where observed image distribution is mapped to the reference image distribution through histogram equalisation, is conducted on each eye fundus image to acquire similar colour appearance over the image set. If $p_r(\nu)$ is the reference images's probability density function with respect to an intensity value ν then the mapping $T_r(\nu)$ defines the histogram equalisation as follows:

$$T_r(\nu) = \int_{\nu} p_r(\nu) d\nu. \quad (5.20)$$

If the mapping $T_s(\nu)$ for observed image is computed similarly then the intensity value ν can be mapped to the reference image distribution by finding the inverse transform

$\hat{\nu} = T_r^{-1}(T_s(\nu))$ [67]. In the case of colour images the procedure is conducted on each colour channel separately.

To compensate the unwanted artefacts caused by the histogram matching, the kernel density estimation is used to smooth the image channel distributions [67, 148]. For example, the probability density function $p_r(\nu)$ can be approximated as a sum of Gaussians centred on each pixel value ν_i in the image as follows:

$$p_r(\nu) = \frac{1}{n} \sum_{i=1}^n \mathcal{N}(\nu | \nu_i, \sigma^2), \quad (5.21)$$

where n is the pixel count in the image and σ is the smoothing factor (bandwidth). By approximating both distributions the smoothing factor becomes a trade-off factor between the distributions. A value of $\sigma = 0.05$ was empirically chosen. Example histogram matching is shown in Fig. 5.5.

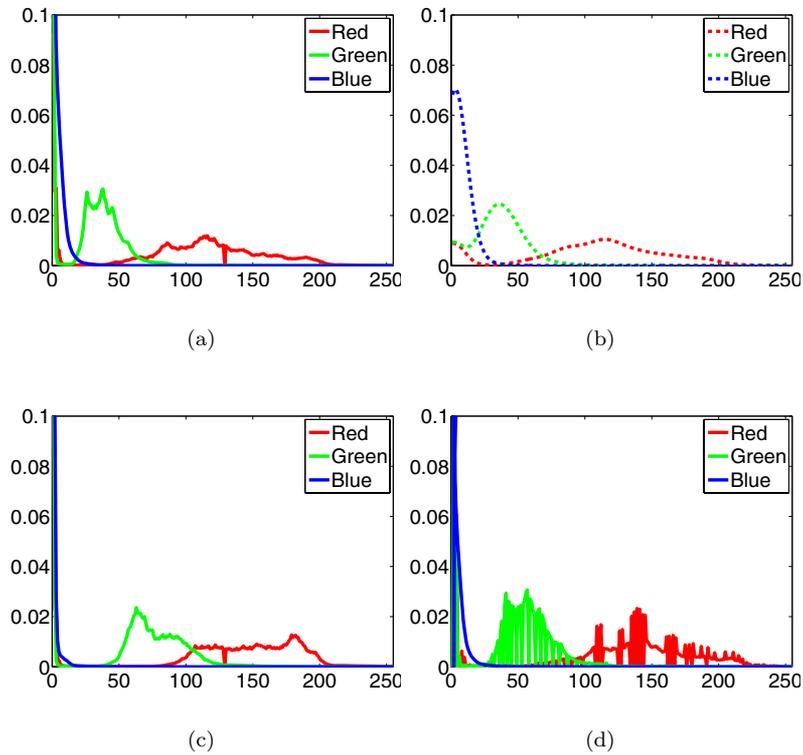


Figure 5.5: Colour correction based on histogram matching: (a) reference image histogram; (b) kernel density smoothed reference image histogram; (c) observed image histogram; (d) corrected image histogram.

5.7.2 Colour transfer

Colour transfer is a straightforward colour correction method to transfer colour characteristics of an image to another based on simple image statistics [134]. First, source and target images are transformed into colour space which minimises the colour channel correlations, and then the mean and standard deviations of the two images are matched.

In the construction of decorrelated colour space, an ensemble of spectral images that represented naturally occurring images were converted into LMS colour space, where L, M, and S corresponds to the spectral sensitivities of human cones [137]. The principal component analysis was applied to LMS colour data to maximise the decorrelation between axes. As a result, for given point in the LMS colour space the following transform were suggested [134]:

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \log L \\ \log M \\ \log S \end{bmatrix}, \quad (5.22)$$

where l represents an achromatic channel and α and β are the chromatic yellow – blue and red – green opponent channels. The LMS values are converted from RGB values via XYZ colour space using device independent conversion that maps the white in the chromaticity diagram (CIE xy) to white in the RGB space. Since the white in the chromaticity diagram is defined as $x = X/(X + Y + Z) = 0.333$ and $y = Y/(X + Y + Z) = 0.333$, the required conversion maps $X = Y = Z = 1$ to $R = G = B = 1$. In consequence, the desired RGB-to-LMS mapping becomes [134]

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (5.23)$$

When the source and target image are mapped to $l\alpha\beta$ colour space, the colour transfer is conducted by matching the means and standard deviations of each axis separately. For each component in the $l\alpha\beta$ colour space the corrected values can be computed as follows:

$$\hat{c} = \frac{\sigma^t}{\sigma^s}(c^s - \langle c^s \rangle) + \langle c^t \rangle, \quad (5.24)$$

where c^s is the component pixel value in the source image, \hat{c} is the corrected pixel value, $\langle c^s \rangle$ is the mean, and σ^s is the standard deviation. The corresponding mean and standard deviation for the target image are $\langle c^t \rangle$ and σ^t . The correction is conducted component-wise for for all the pixels in the source image. The corrected image in $l\alpha\beta$ colour space is converted back to RGB via LMS and XYZ colour spaces using the following transformations:

$$\begin{bmatrix} \log L \\ \log M \\ \log S \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{6} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix}, \quad (5.25)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.4679 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}. \quad (5.26)$$

5.7.3 Geometric colour correction using image landmarks

Geometric colour correction using landmarks utilises the colour information of visible image landmarks such as optic disc, fovea, blood vessels, fundus, and the black area outside the fundus. The idea is that a landmark set defined by the mean colour of each landmark characterises the image colour content (Fig. 5.6). To normalise the colour appearance over the image set, a common reference landmark set is constructed, where the images (landmark sets) are mapped using linear transformations. In the literature, a similar method is used for object localisation [82].

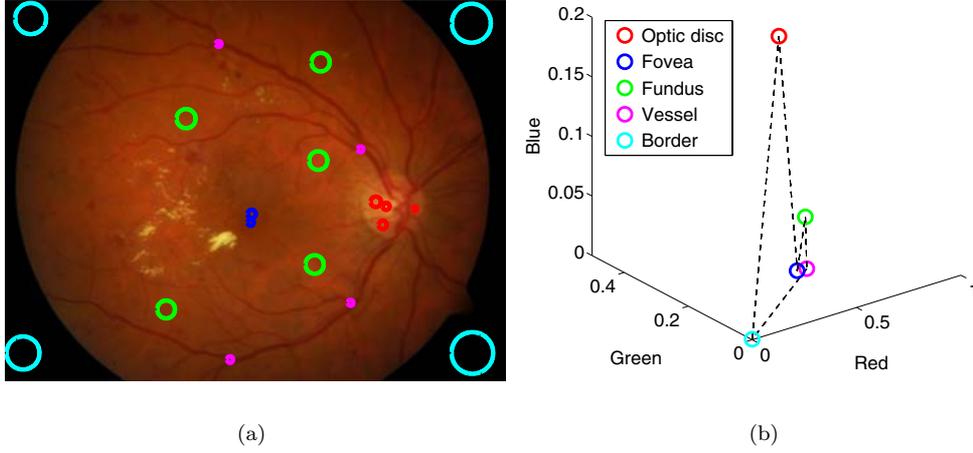


Figure 5.6: Example landmark set: (a) Manually annotated areas of retinal landmarks in an eye fundus image (red=optic disc, green=eye fundus, magenta=blood vessels, blue=fovea and cyan=border); (b) Landmarks set defined by the mean colour of the annotated landmark areas.

A landmark set of n landmarks with mean colours (r_i, g_i, b_i) is represented as a vector $\mathbf{X} = [r_1, \dots, r_n, g_1, \dots, g_n, b_1, \dots, b_n]^T$. For given K training landmark sets (Fig. 5.7), the reference landmark set is constructed by aligning each training set so that the sum distances of each training set to the reference is minimised:

$$\epsilon = \sum_{i=1}^K |\bar{\mathbf{X}} - T_i(\mathbf{X}_i)|^2, \quad (5.27)$$

where $\bar{\mathbf{X}}$ is the reference landmark set, \mathbf{X}_i is the i th example landmark set and T_i is the alignment operator (linear transformation) for i th example landmark set.

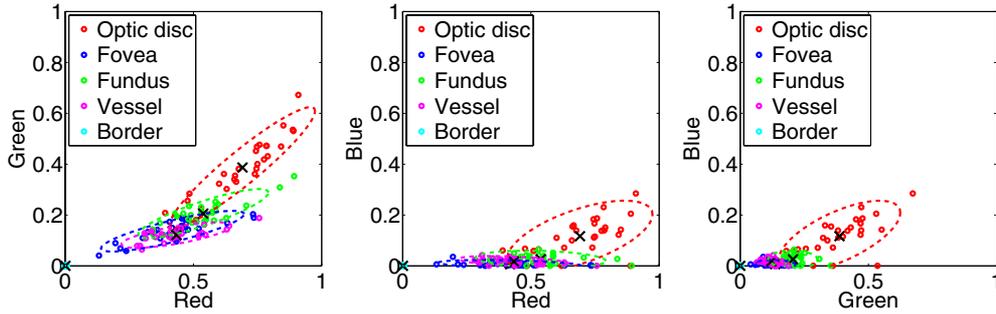


Figure 5.7: Training landmark sets extracted from the manually annotated areas in the training images and plotted to the RGB colour space (dashed line denotes the landmark variance and the black “x” denotes the mean).

To solve the reference landmark set in Eq. 5.27, an iterative solution is proposed based on the algorithm in [24]. A random training landmark set is selected as the initial reference set. Each training set, at a time, is selected and transformed to the reference set by estimating the corresponding homography. After each transformation, the reference set is updated by computing the average landmark set from the transformed training sets. These procedures are repeated until each training set is used. Several random seeds are tried and the best is selected as the final reference landmark set. The reference landmark set construction is presented in Algorithm 3.

Algorithm 3 Reference landmark set construction

```

1: for Number of random seeds do
2:   Current reference set (curr_ref_set)  $\leftarrow$  Select a random training landmark set
3:   for each training landmark set (curr_train_set) do
4:     Estimate homography from the curr_train_set to the curr_ref_set
5:     Transform the curr_train_set to the curr_ref_set
6:     curr_ref_set  $\leftarrow$  mean of transformed training landmark sets
7:   end for
8:   Transform all the training landmark sets to the curr_ref_set
9:   Compute the sum distances ( $\epsilon$ ) of each transformed training set to the curr_ref_set
10:  if  $\epsilon < \epsilon_{best}$  then
11:    best_ref_set  $\leftarrow$  curr_ref_set
12:  end if
13: end for
14: Return the best reference landmark set (best_ref_set)

```

The colour correction step for an image is straightforward: 1) extract and compute the landmark set; 2) estimate linear transformation from the landmark set to the reference set; 3) colour correct the image using the estimated linear transformation. The main motivation for using colour correction in this thesis is to study on how the colour correction

affects the lesion detection, and therefore the images are corrected using landmark sets derived from the manually annotated image areas. An example correction is illustrated in Fig. 5.8.

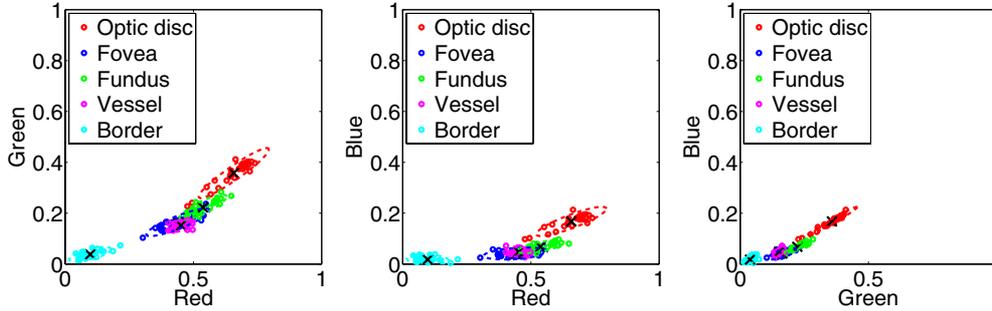


Figure 5.8: Training landmark sets illustrated in Fig. 5.7 are transformed to the reference landmark set using the similarity transform (black “x” denotes the reference landmark locations and the dashed line denotes the variance of transformed training landmarks).

5.7.4 Experimental results and discussion

The quantitative DIARETDB1 results for the colour correction are shown in Tables 5.7 and 5.8 (Entry 1). The mean ROC curves are visualised in Figs. B.4 and C.4 in Appendices B and C. The quantitative DIARETDB1 results for the illuminance and colour correction are given in the same tables (Entry 2) and the mean ROC curves are visualised in Figs. B.5 and C.5 in Appendices B and C. In the tables, Geometric 1 corresponds to a linear transformation under similarity using the following landmarks: blood vessels, eye fundus and black background. Geometric 2 corresponds to the same transformation, but the following landmarks are used: optic disc, fovea, blood vessels, fundus, and black background.

Colour correction methods clearly transform the appearance of eye fundus images to a more similar representation (Figs G.1 and H.1 in Appendices G and H). However, the quantitative results show that histogram matching is the only method that essentially improved the results. Since the histogram matching is a non-linear operation, the use of linear transformations in the geometric colour correction is not well-grounded. In addition, colour correction on its own performed better than the combination of illuminance and colour correction. This can be explained by the fact that illuminance correction is partially included in the colour correction methods and the used feature (pixel colour) did not exploit the spatial information (e.g. texture).

In general, the diabetic retinopathy seems to affect the colour of the retina as a whole due to its metabolic nature. By normalising the images to incorporate the same colour content, colour information between normal and abnormal images becomes mixed. The confusion of the underlying colour distributions of normal and abnormal images is not a desired property from the diagnostic point of view. Therefore, the methods presented

are only recommended for visualisation purposes, although the quantitative results imply the colour correction can be useful in the case of diabetic retinopathy.

Table 5.7: Mean EER results (image-based evaluation) for the baseline algorithm using colour correction. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

		Ha	MA	HE	SE	Avg.
	Baseline (RGB)	0.29	0.48	0.24	0.25	0.31
Entry 1	Colour transfer	0.66	0.54	0.53	0.58	0.58
	Geometric 1	0.34	0.56	0.38	0.33	0.41
	Geometric 2	0.43	0.53	0.37	0.46	0.45
	Histogram match.	0.29	0.41	0.20	0.20	0.27
Entry 2	Poly. (ord. 4) + Col. Trans.	0.34	0.46	0.30	0.50	0.40
	Poly. (ord. 4) + Geometric 1	0.49	0.48	0.43	0.39	0.45
	Poly. (ord. 4) + Geometric 2	0.40	0.51	0.38	0.54	0.46
	Poly. (ord. 4) + Hist. match.	0.33	0.41	0.20	0.28	0.31

Table 5.8: Mean EER (pixel-based evaluation) results for baseline algorithm using colour correction. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

		Ha	MA	HE	SE	Avg.
	Baseline (RGB)	0.37	0.36	0.39	0.41	0.38
Entry 1	Colour transfer	0.49	0.37	0.39	0.40	0.41
	Geometric 1	0.36	0.33	0.32	0.39	0.35
	Geometric 2	0.37	0.36	0.36	0.44	0.38
	Histogram match.	0.38	0.32	0.30	0.34	0.34
Entry 2	Poly. (ord. 4) + Col. Trans.	0.37	0.36	0.37	0.39	0.37
	Poly. (ord. 4) + Geometric 1	0.40	0.35	0.37	0.45	0.39
	Poly. (ord. 4) + Geometric 2	0.37	0.38	0.38	0.43	0.39
	Poly. (ord. 4) + Hist. match.	0.35	0.36	0.31	0.39	0.35

5.8 Applicability of the image analysis results

Previously, the use of photometric information in the detection of diabetic retinopathy was studied using images that were captured with single eye fundus camera. The photometric cue of the lesions were given in the form of trichromatic RGB values. Since RGB values are dependent on the imaging device, it was reasonable to question the applicability of the results, that is, what if the steps were performed on images that were captured with different eye fundus camera or with multiple eye fundus cameras. To answer these questions, additional experiments were conducted using an independent non-public eye fundus image database containing 107 fundus images with ground truth for exudates that correspond to hard exudates in DIARETDB1. The database is referred as BRISTOLDB and a detailed description was given in Section 3.7.3.

In the first set of experiments (Set 1), the baseline algorithm, illuminance correction (polynomial), colour correction (histogram matching), and their combinations were tested

on BRISTOLDB database similarly as for the DIARETDB1 database. The idea of the first set was to study whether the DIARETDB1 results are specific for the database or not. In the second set of experiments (Set 2), the baseline algorithm was run on the BRISTOLDB test set and trained with the DIARETDB1 train set, where the aim was to maximise the photometric variation between the train set and the test set, and to study whether the colour correction can compensate that variation.

5.8.1 Results and discussion

The quantitative BRISTOLDB results are shown in the Table 5.9 and the mean ROC curves for the Set 1 are visualised in Fig. 5.9. The quantitative results (Set 1) state that BRISTOLDB are not fully in accordance with the DIARETDB1 results. In the case of DIARETDB1, illuminance and colour correction improved the results, whereas the pixel- and image-based results for the BRISTOLDB are contradictory. In addition, the image-based results for DIARETDB1 are clearly better than the corresponding BRISTOLDB results. The differences can be explained by the fact that the expert provided ground truth of DIARETDB1 notably differs from BRISTOLDB. The expert annotations in the BRISTOLDB are restricted to the clearly visible hard exudate areas, whereas the DIARETDB1 annotations are less precise including the lesion affected surrounding areas. As a result, the databases provide two different photometric representations for hard exudates that affect the training and most importantly the evaluation. Thus, one should not expect one-to-one results between the databases, and only general conclusions about the applicability of the DIARETDB1 results can be drawn.

The contradiction between the pixel- and image-based results of BRISTOLDB (Set 1) was caused by the false positive non-exudate pixels that resembled the hard exudate colour. Since the non-exudate pixels were DR related and appeared in the same images as the hard exudates, only the pixel-based results were affected. By using colour correction, the non-exudate pixels were normalised to correspond to the normal background colour, but as a trade-off the optic disc and hard exudate colour distributions became overlapped. Since each image contained high likelihoods in the optic disc region, the image-based results were degraded. This result, however, supports the discussion in Sec. 5.7. In general, illuminance correction did not seem to be essential for neither of the databases, whereas colour correction seems to produce undesired by-products.

The quantitative results (Set 2) address that two uncalibrated eye fundus cameras map the lesion distributions in two different locations in the RGB colour space that cannot be avoided by using the presented colour correction methods. In other words, the colour distribution of hard exudates in the DIARETDB1 training images does not correspond to hard exudates in the BRISTOLDB images even if colour correction is used. It follows that high likelihoods are given for incorrect pixels, and unpredictable results are produced. As a conclusion for the Set 2 of experiments, the photometric information of multiple eye fundus cameras should not be mixed without colour calibrating the imaging systems first.

Table 5.9: Mean EER results for the baseline algorithm using illuminance and colour correction, and their combination in hard exudate detection. The results are based on experiments with DIARETDB1 (DDB1) and BRISTOLDB (BDB) databases.

		Image-based		Pixel-based	
		BDB	DDB1	BDB	DDB1
Set 1	Baseline (RGB)	0.35	0.24	0.34	0.39
	Polynomial (ord. 4)	0.30	0.25	0.36	0.33
	Histogram matching	0.42	0.20	0.34	0.30
	Poly. (ord. 4) + Hist. match.	0.35	0.20	0.37	0.31
Set 2	Baseline (RGB)	0.46	0.24	0.63	0.39
	Colour transfer	0.48	0.53	0.43	0.39
	Geometric 1	0.46	0.38	0.55	0.39
	Geometric 2	0.49	0.37	0.45	0.36
	Histogram matching	0.49	0.20	0.49	0.30
	Poly. (ord. 4) + Col. trans	0.53	0.30	0.48	0.37
	Poly. (ord. 4) + Geometric 1	0.49	0.43	0.55	0.37
	Poly. (ord. 4) + Geometric 2	0.49	0.38	0.49	0.38
Poly. (ord. 4) + Hist. match.	0.40	0.20	0.45	0.31	

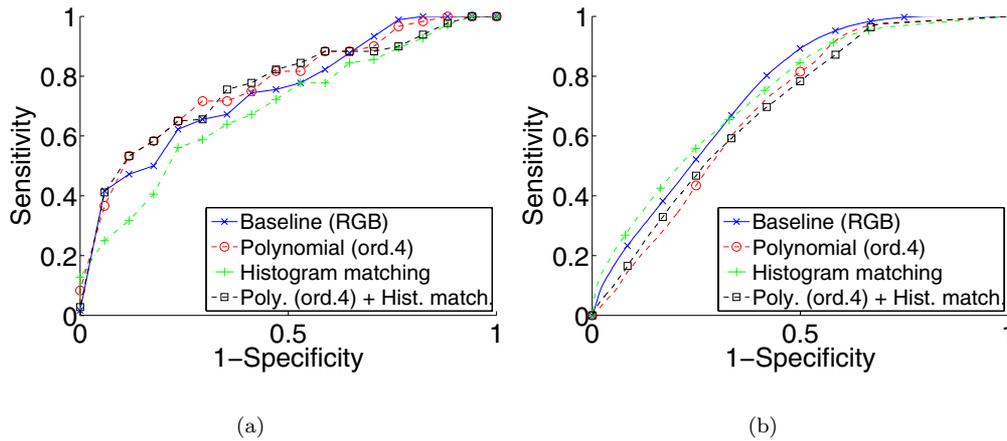


Figure 5.9: Mean EER results for the baseline algorithm using illuminance and colour correction, and their combination in hard exudate detection. The results are based on experiments with BRISTOLDB database: (a) Image-based evaluation; (b) Pixel-based evaluation.

5.9 Under-and overexposed pixels in eye fundus images

There are also other issues related to all images, such as saturated pixels (due to overexposure) or low signal pixels (due to underexposure) which are sources of errors. The pixel is overexposed if the incident light at a pixel causes one of the colour channels of the camera sensor to respond at its maximum value. On the other hand, the pixel is underexposed if the incident light at pixel is too weak and blends with the random signal fluctuation of the camera sensor. Colour being the main focus, the amount of over and underexposed pixels in the DIARETDB1 and BRISTOLDB database images and their implications to the photometric cue are analysed. Note that these errors were ignored in the lesion detection steps since the quality of an image remains poor for the automatic analysis in any case.

The channel-wise noise for each image was estimated from the dark pixels outside the eye fundus area using kernel density estimation with Gaussian kernel (Figs. 5.10 and 5.11(a)). The image pixel values were compared to the noise distribution and a pixel was considered to blend noise if the pixel value was less or equal than a threshold. Below the threshold lies 95% percent of the estimated noise distribution. The pixel was considered overexposed if the channel-wise pixel value corresponded to the maximum value. The noise analysis was not possible for BRISTOLDB since the pixels outside the eye fundus area were masked and set to complete black pixels.

The experiments were divided into three categories: 1) how many over and underexposed pixels there were in the images excluding the black border pixels (Fig. 5.11); 2) how many over and underexposed pixels there were in the training data; and 3) how the over and underexposed pixels were divided between the retinal landmarks such as optic disc, fovea, eye fundus background and blood vessels. The experiments were conducted channel-wise using manual annotations.

5.9.1 Results and discussion

The results for the overexposed pixels are given in Table 5.10 and for the underexposed pixels in Table 5.11. Apparently, the guidelines in the image capturing have been different while collecting the databases. The BRISTOLDB database images contained overexposed pixels in the red channel which was clearly allowed in order to produce best possible signal for the blue channel. In the DIARETDB1 database, however, the signal was preserved within the dynamic range of the eye fundus camera that resulted the blue channel pixels to become underexposed. The alarming fact is that every image of BRISTOLDB contained overexposed pixels and every image of the DIARETDB1 contained underexposed pixels. (Tables 5.10 and 5.11). This may address that there is a trade-off between strong red signal and weak blue signal, and the dynamic range of a standard eye fundus camera is not sufficient to record all the required information. As a result, information is lost either due to overexposure or underexposure. Such deficiency would be particularly problematic for the automatic image analysis algorithms which cope poorly with images that contain high noise and saturation levels. Typical overexposed pixels in the eye fundus area for BRISTOLDB image and typical underexposed for DIARETDB1 image is shown in Fig. 5.11. As a conclusion, under- and overexposed pixels have evidently been present while analysing the use of photometric cue in lesion detection, and therefore the results do not represent the full potential but rather the current state.

Table 5.10: Percentages of overexposed pixels in the DIARETDB1 and BRISTOLDB database images. The number of images containing overexposed pixels is denoted bold in the parenthesis.

		DIARETDB1 (21/89)			BRISTOLDB (107/107)		
Channel		Red	Green	Blue	Red	Green	Blue
I	Entire fundus area	0.1	0.0	0.0	1.1	0.1	0.0
II	Haemorrhages	0.1	0.0	0.0	-	-	-
	Microaneurysms	0.0	0.0	0.0	-	-	-
	Hard exudates	0.0	0.0	0.0	7.0	0.9	0.0
	Soft exudates	0.0	0.0	0.0	-	-	-
III	Optic disc	3.0	0.0	0	14.3	7.4	0.0
	Fovea	0.0	0.0	0.0	0.0	0.0	0.0
	Retinal background	0.3	0.0	0.0	0.8	0.0	0.0
	Vessels	0.0	0.0	0.0	0.9	0.0	0.0

Table 5.11: Percentages of underexposed pixels in the DIARETDB1 database images. The number of images containing underexposed pixels is denoted bold in the parenthesis.

DIARETDB1 (89/89)				
Channel		Red	Green	Blue
I	Entire fundus area	0.0	0.1	57.7
II	Haemorrhages	0.0	0.0	72.0
	Microaneurysms	0.0	0.0	72.0
	Hard exudates	0.0	0.0	82.1
	Soft exudates	0.0	0.0	37.4
III	Optic disc	0.0	0.0	8.3
	Fovea	0.0	0.0	92.1
	Retinal background	0.0	0.0	49.5
	Vessels	0.0	0.4	71.6

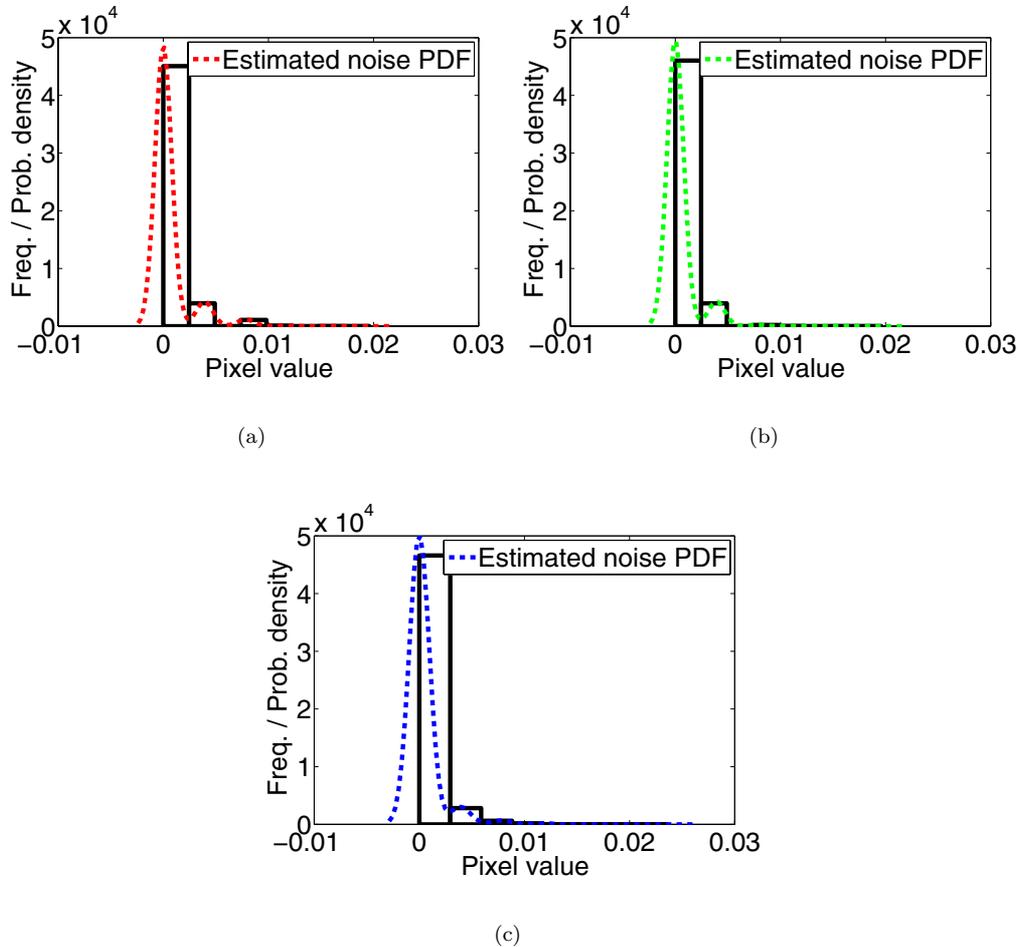


Figure 5.10: Example of channel-wise noise estimation using the kernel density estimation with Gaussian kernel. The noise is estimated from the areas denoted in Fig. 5.11(a). The estimated probability density function (PDF) of noise is denoted with dashed line and plotted over the histogram of the noise pixel values: (a) red image channel; (b) green image channel; (c) blue image channel.

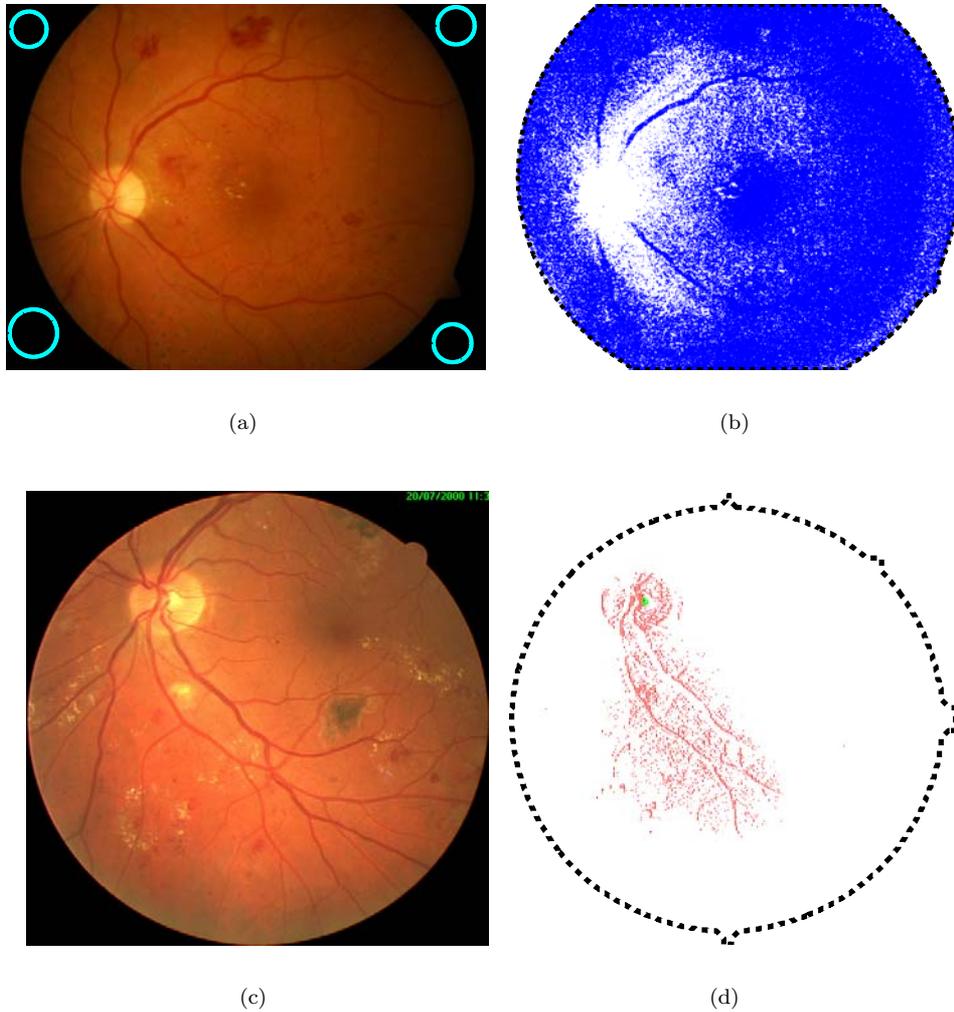


Figure 5.11: Examples of typical under- and overexposed pixels in the DIARETDB1 and BRISTOLDB images (colour denotes the channel of the distorted pixel): (a) example DIARETDB1 image (circles denote the areas used in noise estimation); (b) underexposed pixels blended to the noise in the DIARETDB1 image; (c) example BRISTOLDB image; (d) overexposed pixels in the BRISTOLDB image.

5.10 Conclusion

In this chapter, the use of photometric information in detection of diabetic retinopathy was studied. The following topics were covered: colour space selection, and both illuminance and colour correction. In addition, the background class information to support the lesion detection was discussed. The results are summarised in Table. 5.12. The applicability of the image analysis results was further investigated using an external eye fundus image database (BRISTOLDB) and analysing the amount of over and underexposed pixels involved in the experiments. Based on the analysis and experimental results, it is possible to sketch a photometric cue baseline method for detecting diabetic lesions in eye fundus images solely using photometric information. The photometric cue baseline method is illustrated in Fig. 5.12.

As a conclusion, the background class information, scoring based on probability density likelihoods, and image scoring based on the sum of maxima are essential for the appropriate use of photometric information in detection of diabetic retinopathy. Surprisingly, the illuminance correction and colour correction steps did not seem to be important for successful classification. However, these steps, preferably just one of them, can be performed prior to classification.

Table 5.12: A summary table for the methods used in the experiments. The table comprises mean EER results over the findings. Due to the stochastic nature of GMM-FJ initialisation, training, testing and evaluation of each entry is conducted three times.

Entry	Method	Image-based			Pixel-based		
		min	max	mean	min	max	mean
I	Baseline (RGB)	0.27	0.35	0.31	0.37	0.40	0.38
II	CIE L*a*b*	0.29	0.36	0.32	0.36	0.37	0.36
	HSV	0.32	0.34	0.33	0.34	0.34	0.34
III	Lr100	0.36	0.38	0.37	0.33	0.34	0.34
	Lr1000 (Summax, X = 1%)	0.36	0.38	0.37			
	Lr10000	0.35	0.37	0.36			
	Lr100	0.33	0.36	0.35	0.34	0.37	0.35
	Lr1000 (Summax, X = 80%)	0.33	0.35	0.34			
	Lr10000	0.32	0.33	0.33			
IV	a*b* (CIE L*a*b*)	0.35	0.37	0.36	0.38	0.38	0.38
	rg (R+G+B=1)	0.34	0.34	0.34	0.36	0.37	0.37
	cos ⁴	0.33	0.34	0.34	0.38	0.38	0.38
	Radial poly. (ord. 3)	0.32	0.32	0.32	0.38	0.38	0.38
	Elliptic parab.	0.28	0.35	0.32	0.35	0.36	0.35
	Kang-Weiss (ord. 3)	0.30	0.34	0.31	0.34	0.35	0.35
Polynomial (ord. 4)	0.30	0.31	0.30	0.35	0.37	0.36	
V	Colour transfer	0.53	0.61	0.58	0.41	0.41	0.41
	Geometric 1	0.36	0.43	0.41	0.34	0.36	0.35
	Geometric 2	0.42	0.47	0.45	0.38	0.39	0.38
	Histogram matching	0.27	0.28	0.27	0.33	0.34	0.34
VI	Poly. (ord. 4) + Col. Trans.	0.40	0.40	0.40	0.37	0.37	0.37
	Poly. (ord. 4) + Geometric 1	0.41	0.50	0.45	0.38	0.40	0.39
	Poly. (ord. 4) + Geometric 2	0.40	0.48	0.46	0.39	0.39	0.39
	Poly. (ord. 4) + Hist. match.	0.30	0.31	0.31	0.35	0.35	0.35

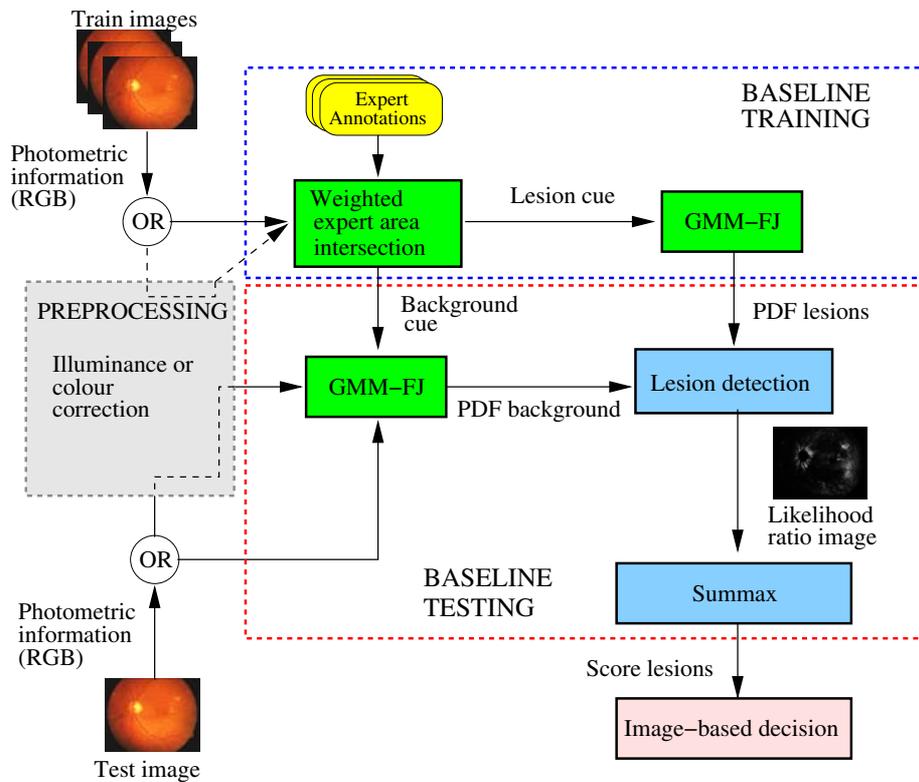


Figure 5.12: Recommended use of the photometric cue for the lesion detection in eye fundus images – the best practise according to this study (the photometric cue baseline method).

Optic Disc Localisation

Normal eye fundus structures, such as blood vessels, vascular arcade, optic disc, macula and fovea, are an essential part in diagnosis of diabetic retinopathy, and fundamental to the subsequent characterisation of the disease. In this chapter, a simple, fast and robust optic disc localisation method using colour decorrelated templates is presented that results in accurate location of the optic disc in colour eye fundus images.

6.1 Introduction

The visible part of the optic nerve in the eye, referred to as the optic disc, is essential when evaluating the existence and severity of diabetic retinopathy. The disease predominantly alters the appearance of the disc, e.g., by developing new fragile blood vessels inside the optic disc region due to the lack of oxygen in the fundus. Observing the state of the optic disc is not only useful in diagnosis of diabetic retinopathy, but also important for diseases, such as glaucoma, diabetic optic neuropathy and other optic nerve related pathologies.

In this chapter, a simple and robust optic disc localisation using colour decorrelated templates is presented. The proposed method captures the colour characteristics of the optic disc at the training stage by using principal component analysis (PCA) to find the vector basis for the colour data extracted from several optic disc regions. The colour decorrelated templates are extracted from the training images projected to the space spanned by the resulting eigenvectors (i.e., the colour decorrelated template space). In the localisation stage, test images are projected to the colour decorrelated template space, where channel-wise template matching using colour decorrelated templates and customised similarity metric is conducted to locate the optic disc. Since optic disc extraction is usually part of a more comprehensive diagnosis system, it is important that the method is efficient and accurately estimates the location of the optic disc. The experiments are conducted using the public fundus image database, DIARETDB1. The content of this chapter is reported in [88].

6.2 Related work

Various methods have been proposed for extracting the optic disc from digital eye fundus images. The round shape and relatively bright appearance of the optic disc are the most commonly exploited features reported in the literature [150, 95, 121, 126]. For example, Sinthanayothin *et al.* assumed that rapid intensity variation inside the optic disc area between the dark blood vessels and bright optic nerve fibres would reveal the optic disc location [150]. Hence, an intensity variance map was computed in a sliding window over an intensity image and the point with maximum variance was chosen as the optic disc location. Park *et al.* determined bright candidate areas using repeated thresholding and measuring the roundness of the thresholded areas [126]. The candidate area contours were estimated with Hough transform, from where the best candidate was selected based on the accumulator value and the average intensity of the circle area.

Alternatively, there are methods that exploit the contextual information, such as blood vessel structure and characteristics [167, 47, 113, 75, 165]. For example, Foracchia *et al.* defined a geometrical model to express the directional behaviour of common vascular network in fundus images using two parabolas and optic disc location as model parameters [47]. The model parameters including the optic disc location were identified using blood vessel centre points and corresponding vessel directions provided by an external vascular network extraction algorithm. Hoover and Goldbaum introduced voting-based localisation method for optic disc referred as fuzzy convergence, where blood vessel segments were modelled by line segments and convergence of the vessel network by line intersections [75]. The optic disc location was defined as the focal point of the fuzzy convergence.

The proposed method utilises template matching and PCA which have been used for optic disc extraction. The related works of significance to our proposed method are the following. Osareh *et al.* performed template matching using normalised cross-correlation on gray level images with a manually generated mean optic disc as a template [121]. This gave an approximate centre for the location of the disc, after which they applied an active contour to approximate the entire disc boundary. Lalonde *et al.* used Hausdorff-based template matching on edge images guided by pyramid decomposition for large scale object tracking [95]. Li and Chutatape [98] applied the PCA based face detection method originally proposed by Turk and Pentland [170] for optic disc extraction. In this chapter, Li and Chutatape's method is used as a comparison method and a more detailed description is given in Section 6.5.

6.3 Pre-processing

Prior to extracting the optic disc location, the colour and non-uniform illuminance of the eye fundus image are corrected, respectively. The decreasing image illuminance towards periphery is corrected using modified version of the method proposed by Grisan *et al.* [64]. First, a parametric model (elliptic paraboloid in Chapter 5.6) is fitted to blood vessel points extracted from an eye fundus image using the method proposed by Soares *et al.* [155]. Then, an image $I_0(x, y)$ affected by the non-uniform illuminance is corrected using an additive correction model as follows:

$$I(x, y) = I_0(x, y) - \phi(x, y) + N_a, \quad (6.1)$$

where $I(x, y)$ is the corrected image, $\phi(x, y)$ is the parametric model fitted to the blood vessel points and N_a defines the correction plane. The colour eye fundus images are transformed from *RGB* into the *HSV* colour space, where the fitting and correction are applied to the value and saturation channels. An example of illumination correction is shown in Fig.6.1. For the colour correction, the previously described histogram matching is used (Sec. 5.7.1).

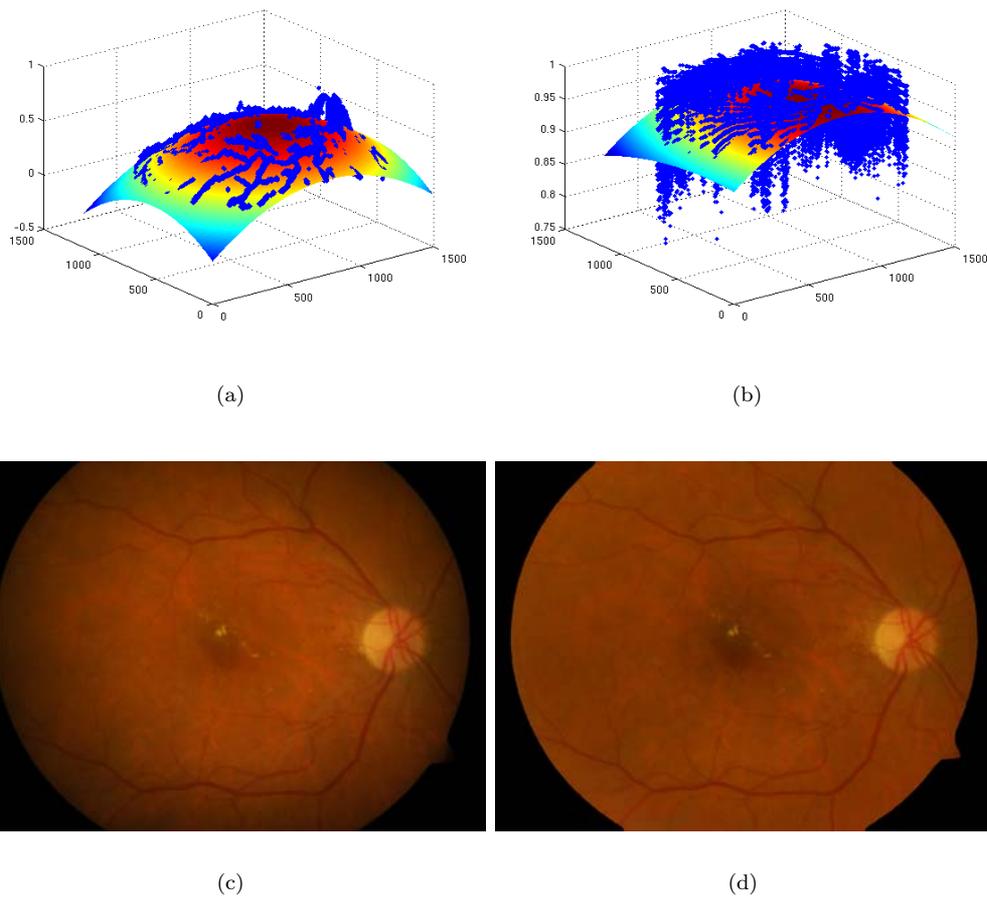


Figure 6.1: Illumination correction example: an elliptic paraboloid fitted to vessel points in a) value channel; b) saturation channel; c) observed image; d) illuminance corrected image.

6.4 Optic disc extraction

The proposed method is described in detail in the following subsections and illustrated in the flowchart in Fig. 6.2.

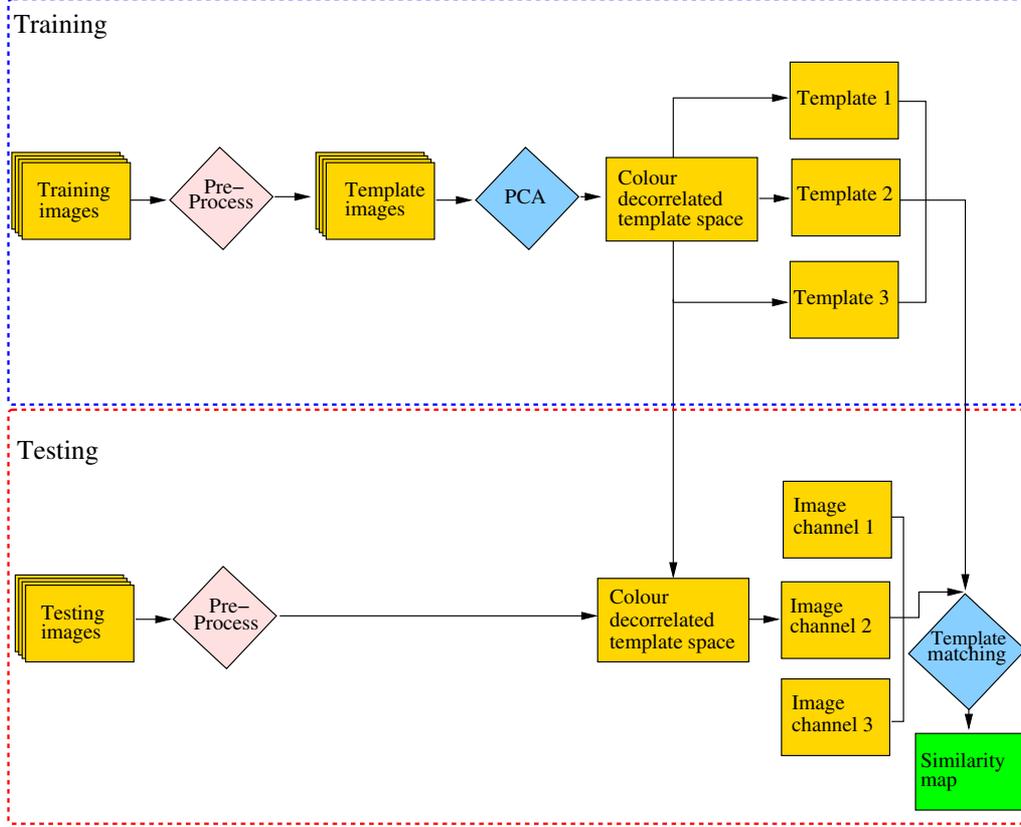


Figure 6.2: Flowchart of the proposed optic disc localisation method.

6.4.1 Colour decorrelated template space

Initially, PCA is performed to determine the vectors that span the optic disc colour basis (colour decorrelated template space) and decorrelate the colour channels. Hence, each channel is independent and describes an optic disc colour characteristic. The method is similar to [192], where colour channel decorrelation was used as part of a random texture analysis application. The colour decorrelated template space is generated from colour data of several optic disc images extracted from the training images.

Let $\mathbf{O} = \{\mathbf{x}_i \in \mathcal{R}^3 | i = 1 \dots n\}$ be the $3 \times n$ matrix containing all the colour pixels, $\mathbf{x}_i = [r_i, g_i, b_i]^T$, of optic disc regions extracted from a set of training images, and $\bar{\mathbf{O}}$ be the mean vector (3×1). Then, Singular value decomposition (SVD) can be used to obtain the eigenvectors $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ of \mathbf{O} which spans the colour decorrelated template space

$\Psi_{O,E}$. The training images are projected to the colour decorrelated template space using the following:

$$\mathbf{Y}' = PCA(\mathbf{Y}, \Psi_{O,E}) = \mathbf{E}^T(\mathbf{Y} - \bar{O}\mathbf{I}), \quad (6.2)$$

where \mathbf{I} is $1 \times m$ unit vector, \mathbf{Y} is $3 \times m$ matrix containing m colour pixels of the image. The colour decorrelated templates are computed from optic disc regions extracted from the resulting training images. Observed image projected to the colour decorrelated template space and colour decorrelated templates is illustrated in Fig. 6.4. As a comparison, the observed image and colour correlated templates are shown in Fig. 6.3.

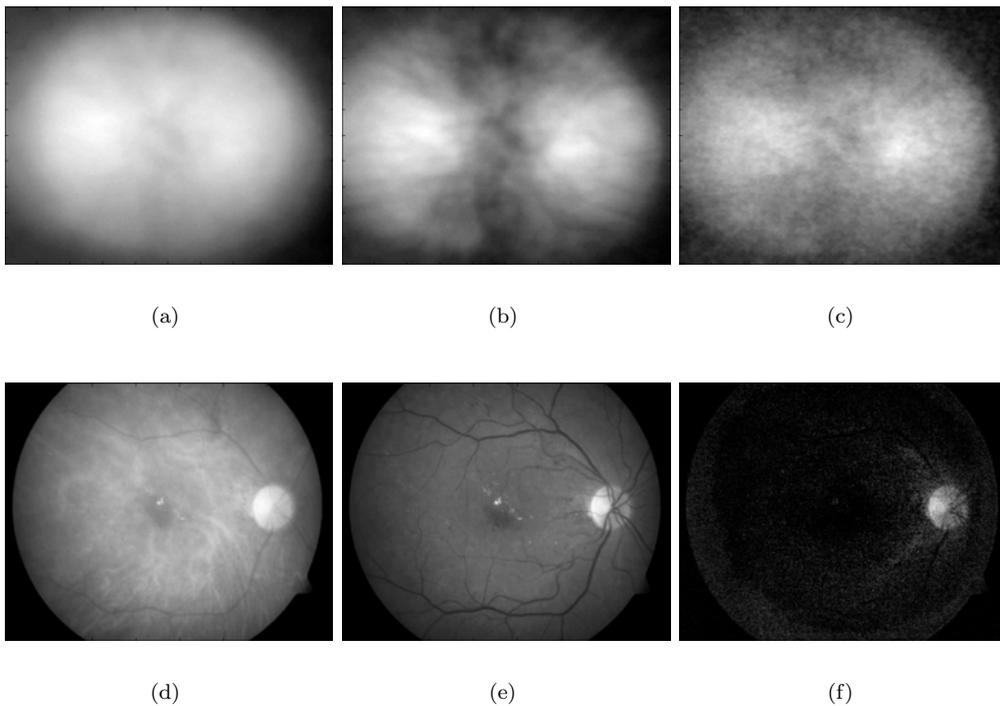


Figure 6.3: Colour correlated templates and observed image: a-c) templates; d-f) image channels.

6.4.2 Template matching

The eigenvectors \mathbf{E} are used to project test images to the colour decorrelated template space before conducting template matching (Fig. 6.4). The similarity between a $N_x \times M_y$ template $t(i, j)$ and the same sized image patch $g(i, j)$ at each image location (x, y) is commonly measured with the squared Euclidean distance:

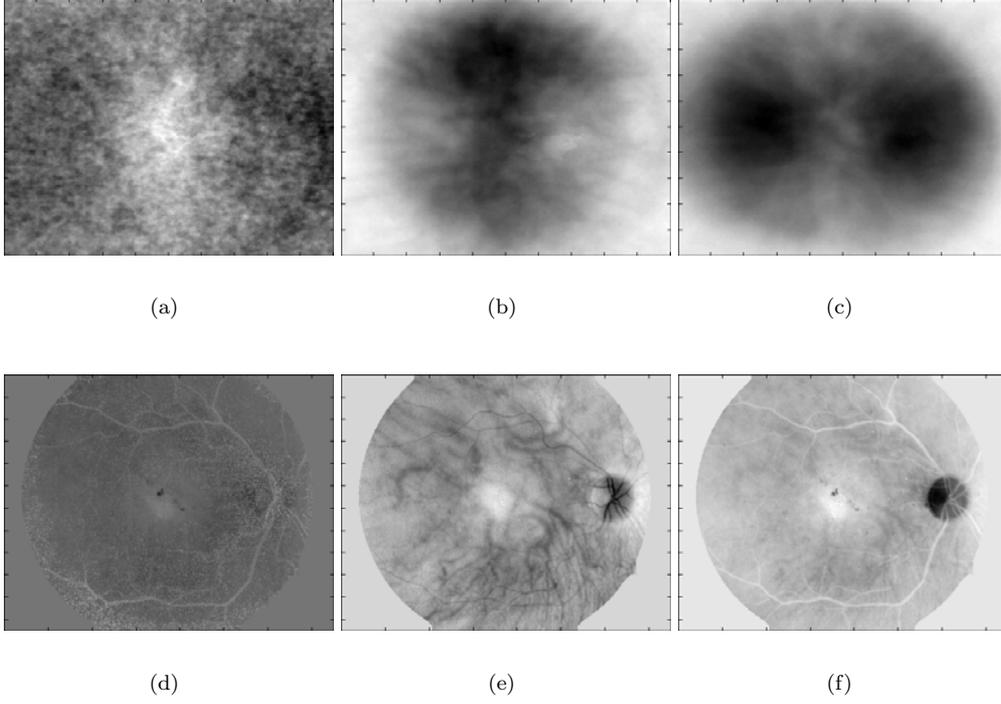


Figure 6.4: Colour decorrelated templates and observed image in colour decorrelated template space: a-c) templates; d-f) image channels.

$$d^2(x, y) = \sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} |g(i, j) - t(i-x, j-y)|^2, \quad (6.3)$$

where $width \times height$ is the image size and the similarity is computed at each image location i.e. $x = 0 \dots width - 1$ and $y = 0 \dots height - 1$. Alternatively, a computationally more efficient cross-correlation motivated by the squared Euclidian distance can be used:

$$CC(x, y) = \sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} g(i, j)t(i-x, j-y). \quad (6.4)$$

The cross-correlation over the image can be computed efficiently via the fast Fourier transform, as used here. The cross correlation assumes small illumination variations over the image which is not valid in eye fundus images. In such cases the normalised cross-correlation (NCC) is typically applied [164]:

$$NCC(x, y) = \frac{CC(x, y)}{\sqrt{\sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} g(i, j)^2} \sqrt{\sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} t(i-x, j-y)^2}} \quad (6.5)$$

However, due to the bright appearance of the optic disc in the fundus images, the denominator term $\sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} g(i, j)^2$ in Eq. 6.5 appears to have high responses on correct matching locations. Moreover, the feature is preserved in the transformation to colour decorrelated template space (Fig. 6.5). Instead of using NCC , the presented optic disc extraction method compensates the uneven image illuminance in the pre-processing and favours bright areas in the template matching by defining a customised similarity metric $\lambda(x, y)$:

$$\lambda(x, y) = \frac{CC(x, y)}{\sqrt{\sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} t(i-x, j-y)^2}} \quad (6.6)$$

This form of template matching is performed on each channel separately in the colour decorrelated template space. The term $\sqrt{\sum_{i=x}^{x+N_x-1} \sum_{j=y}^{y+M_y-1} t(i-x, j-y)^2}$ in Eq. 6.6 is independent of image location (m, n) and removes the emphasis of the template in each channel. The final indicator of optic disc location in the fundus image is the sum of the results from the individual channels, resulting in the similarity map $s(x, y) = \sum_e \lambda(x, y)$, where e is the channel. The maximum of the similarity map $s(x, y)$ indicates the optic disc location (x, y) (Fig. 6.6).

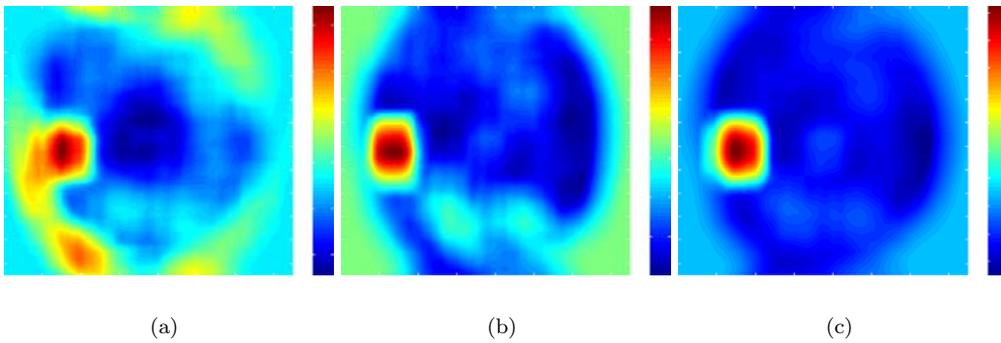


Figure 6.5: The denominator term value at each image location in colour decorrelated template space: a) channel 1; b) channel 2; c) channel 3.

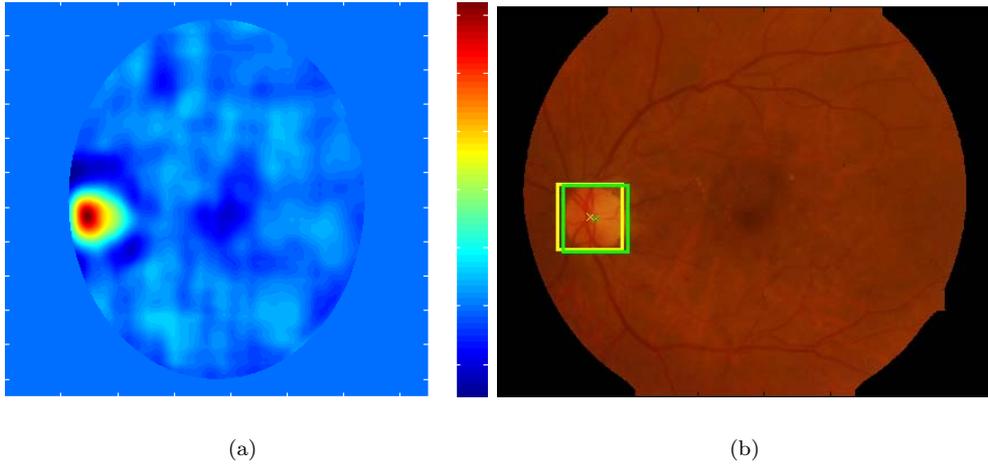


Figure 6.6: Example optic disc extraction result: a) similarity map $s(x, y)$; b) example extraction (ground truth in yellow).

6.5 Experiments

Li and Chutatape also proposed a PCA-based optic disc extraction method that was used as a reference method in the experiments [98]. Since the methodology used in the Li and Chutatape’s method is well-known in computer vision and especially efficient in face recognition [171], the method was considered suitable for comparative analysis. The decision was reinforced by the fact that Li and Chutatape reported successful results with all of their test images (40 images).

In Li and Chutatape’s method, the PCA was applied to the optic disc regions extracted from the training set of gray level images to determine the features that characterise the variations of the optic disc appearance [98]. These linearly independent features were referred to as eigendiscs. The method assumed that each input optic disc image can be approximated as a linear combination of the eigendiscs and the mean training set optic disc. If the squared Euclidean distance between a test input image and the approximated test image was small the input image is considered as an optic disc. The squared Euclidean distance was computed in a sliding window on every image point and the point with the minimum distance was chosen to be the optic disc location. In practice, Li and Chutatape computed the Euclidean distance only on bright candidate areas determined in a pre-processing stage to reduce the computational load.

The manually annotated optic disc locations of the DIARETDB1 database were used for the comparative analysis. The proposed method was trained using the training set of 28 images defined in the DIARETDB1 database. Since the optic disc localisation was considered as a preliminary step for a more comprehensive system, the accuracies of the proposed method and Li and Chutatape’s method were evaluated using the entire database. The mean optic disc diameter in the training set images was approximately

200 pixels, therefore a 200×200 template was used for the evaluation basis. A threshold distance from the ground truth was varied and the accuracy was the percentage of detected optic discs inside the threshold distance. The distance from ground truth was normalised with the length of optic disc radius (100 pixels). Evaluation results of the colour decorrelated templates are given in Fig. 6.7. The colour decorrelated templates with pre-processing and Li and Chutatape's method were compared and the results are shown in Fig. 6.8. Example computation times for colour decorrelated templates and Li and Chutatape's method are given in Table 6.1.

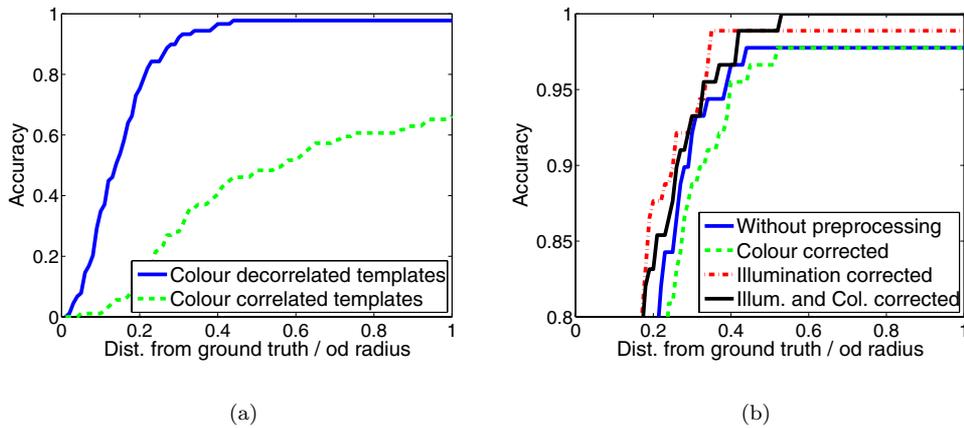


Figure 6.7: Evaluation results of the colour decorrelated templates: a) colour decorrelated templates without pre-processing; b) colour decorrelated templates with pre-processing.

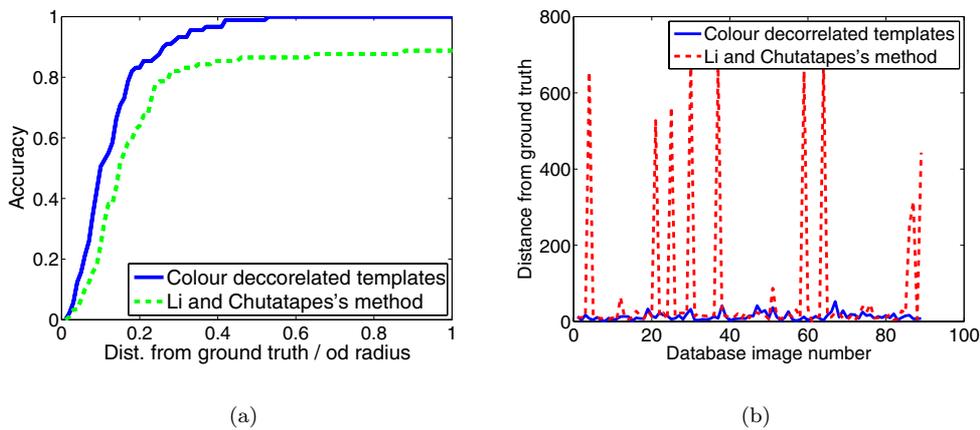


Figure 6.8: Evaluation results of the comparative analysis: a) Li and Chutatape's method; b) distance from ground truth.

Table 6.1: Computation time evaluation results for the ten first database images and distances between detected optic discs and ground truths in pixels. Distances over optic radius (100 pixels) marked with black.

	Colour decorrelated templates		Li and Chutatape’s method	
	Distance from ground truth (pixels)	Computation time(s)	Distance from ground truth (pixels)	Computation time(s)
image001	9.4868	9.3874	10.1980	21.6208
image002	6.7082	8.9209	10.4403	18.2414
image003	15.5242	8.6041	11.1803	19.0500
image004	15.5242	8.6787	654.7648	33.3511
image005	2.0000	8.4736	14.0357	18.7710
image006	14.0000	8.5706	10.7703	18.8342
image007	2.2361	8.4846	8.0623	26.9084
image008	2.8284	8.7475	9.0554	18.0940
image009	8.0623	8.6124	6.0000	18.4484
image010	2.2361	8.531	10.0000	18.8954
Average	7.8606	8.7011	74.4507	21.2215
Accuracy	100%		90%	

6.6 Discussion

Li and Chutatape’s method determined candidate areas in the pre-processing using pixels with the highest 1% gray levels in the intensity image and then removing areas with pixels under a predefined threshold. This restricted the minimum squared Euclidean distance search space and pruned some false optic disc locations. On the other hand, the pre-processing steps caused their method to miss dark optic discs in the periphery of the eye fundus area resulting high error distances in Fig. 6.8(b) and Table 6.1.

Computationally, the performance of the colour decorrelated templates was extremely fast. The pre-processing time was not included as it is essential for any further analysis of disease detection, such as exudates [121]. The computation of colour decorrelated template space and template matching with colour decorrelated templates via FFT were very fast to compute and easy to implement. Compared to Li and Chutatape’s method, the computation of the colour decorrelated templates over image involved 3 correlations, whereas their method involved $K + 2$ correlations using the formulation by Turk and Pentland [170]. K is the number of eigendiscs (eigenvectors). In the experiments, 8 eigendiscs out of 28 were used for Li and Chutatape’s method based on the eigenvalues. Computation times are given in Table 6.1 for the first ten database images.

6.7 Conclusion

In this chapter, a simple, robust and automated optic localisation method using colour decorrelated templates was presented for digital colour fundus images. The experimental

results using the public fundus image database DIARETDB1 showed that the proposed method is robust and results in an accurate optic disc location.

There were three main issues that were addressed in this thesis: 1) how to utilise the photometric cue in the detection of diabetic retinopathy, 2) how good is the photometric cue in the detection of diabetic retinopathy, and 3) how the performance of the diabetic retinopathy detection algorithms should be evaluated to produce reliable comparison. In addition, a simple, robust and automated optic localisation method using colour decorrelated templates was presented as a supplementary item.

Public image databases are an essential resource in the development and comparison of eye fundus image analysis algorithms, and therefore a framework for constructing benchmarking databases was presented in Chapter 3. The framework provides guidelines on how to construct an image database that comprise true patient images, ground truth and an evaluation protocol. Since standard receiver operating characteristics (ROC) analysis is a frequently used tool to assess algorithm performance and it is well in accordance with the diagnostic procedures in medical practice, it was adopted in the framework to provide means for statistical performance evaluation. Hence, the framework presented pixel- and image-based evaluation protocols for assessing the detection performances of eye fundus image analysis algorithms. The essential information on how the true patient images should be collected, stored, annotated and distributed, was assembled while collecting the eye fundus image databases, DIARETDB0 and DIARETDB1. To promote automatic detection of diabetic retinopathy and help other researches to collect and publish their data, the collected databases were put publicly available including the expert annotations. In addition, reusable tools will be available for image annotation, algorithm evaluation, and ground truth estimation

Colour is an important cue to detect abnormalities in eye fundus images, and since the eye fundus images are almost without exception either monochromatic (gray level) or trichromatic colour images, it was important to study the goodness of the colour cue in the detection of diabetic retinopathy. For the purpose, a supervised algorithm (baseline algorithm) based on one-class classification and Gaussian mixture model estimation was presented in Chapter 3. The presented algorithm distinguishes a diabetic lesion type from all other possible objects in eye fundus images by modelling only the colour cue of

that certain lesion type. In the training, the algorithm combines manual annotations of several experts to generate the image locations for training the one-class classifier. The training strategy was selected experimentally in Chapter 4, where the use of multiple expert information in ground truth estimation and classifier training was discussed. Finally, the goodness of the photometric cue in the detection of diabetic retinopathy was quantitatively evaluated in Chapter 5 by assessing the baseline algorithm’s performance using the benchmarking framework while conducting experiments with the colour space selection, both illuminance and colour correction, and background class information. The applicability of the image analysis results were further studied by conducting the same experiments using an external eye fundus image database (BRISTOLDB0) and by analysing the amount of over and underexposed pixels involved in the experiments. The best methods are summarised in Tables 7.1 and 7.2.

Table 7.1: Summary of the baseline algorithm results using the DIARETDB1 database. The table describes the sensitivity and specificity for the ROC curve point, where the sensitivity is equal to specificity, i.e. table value = sensitivity = specificity..

	HA	MA	HE	SE	Avg
Task: detect image pixels with diabetic lesion type					
Colour space selection (RGB)	0.63	0.64	0.61	0.59	0.62
Counter-example information (Lr10000, $X = 80\%$)	0.65	0.65	0.73	0.63	0.66
Illuminance correction (Polynomial, ord. 4)	0.65	0.64	0.67	0.60	0.64
Colour correction (Histogram matching)	0.62	0.68	0.70	0.66	0.66
Col. and illum. correction (Poly. + Hist.match.)	0.65	0.64	0.69	0.61	0.65
Task: detect images with diabetic lesion type					
Colour space selection (RGB)	0.71	0.52	0.76	0.75	0.69
Counter-example information (Lr10000, $X = 80\%$)	0.69	0.51	0.76	0.72	0.67
Illuminance correction (Polynomial ord. 4)	0.67	0.65	0.75	0.72	0.70
Colour correction (Histogram matching)	0.71	0.59	0.80	0.80	0.73
Col. and illum. correction (Poly. + Hist. match.)	0.67	0.59	0.80	0.72	0.69

Table 7.2: Summary of the baseline algorithm results using the BRISTOLDB database. The table describes the sensitivity and specificity using a ROC curve point, where the sensitivity is equal to specificity, i.e. table value = sensitivity = specificity.

Hard exudates (HE)	
Task: detect image pixels with diabetic lesion type	
Colour space selection (RGB)	0.66
Illuminance correction (Polynomial, ord. 4)	0.64
Colour correction (Histogram matching)	0.66
Col. and illum. correction (Poly. + Hist. match.)	0.63
Task: detect images with diabetic lesion type	
Colour space selection (RGB)	0.65
Illuminance correction (Polynomial, ord. 4)	0.70
Colour correction (Histogram matching)	0.58
Col. and illum. correction (Poly. + Hist. match.)	0.65

Based on the DIARETDB1 and BRISTOLDB results, it seems that diabetic retinopathy affects the colour of the retina as a whole, and forcing the same colour appearance using colour correction can be harmful for the colour distributions of normal and diseased images. Secondly, the colour space and non-uniform illuminance correction did not seem to be essential for the successful classification. However, if the used colour feature had exploited spatial dependencies, i.e. the information in the neighbouring pixels, the non-uniform illuminance correction would most likely have proven to be useful. In general, the illuminance correction is task specific and the general approaches presented in this thesis are not the best possible in the case of eye fundus images. Otherwise, the background class information, scoring based on probability density likelihoods, and image scoring based on the sum of maxima are essential when using photometric cue in the lesion detection.

To determine the goodness of the photometric cue in the detection of diabetic retinopathy, the baseline algorithm results are compared to the detection performances reported in the literature. This is important even though different data sets were used in the development. From the literature review in Section 2.5, the approaches that produced the best results in the detection of haemorrhages, microaneurysms, hard and soft exudates were selected for the comparison. The methods that used colour eye fundus images or the DIARETDB1 database in the experiments were prioritised over other approaches. Moreover, the “red lesion” and “bright lesion” detection algorithms were excluded from the selection. The reported performances of the selected state-of-the-art approaches are shown in Table 7.3. Note that none of the reviewed methods reported pixel-level performance for haemorrhage or soft exudate detection, and therefore the corresponding table items are left blank. Based on the above-mentioned conclusions regarding the use of photometric information, the performances of the state-of-the-art methods are compared to the baseline algorithm results in Table 7.1 denoted with “counter-example information (Lr10000, $X = 80\%$)”.

Table 7.3: Reported performances of the state-of-the-art approaches that were selected for the method comparison, where SN = sensitivity, SP = specificity and PPV = positive predictive value.

	SN	SP	PPV
<u>Task: detect image pixels with diabetic lesion type</u>			
Haemorrhages	-	-	-
Microaneurysms (Bhalerao et al. [20])	0.83	0.80	-
Hard exudates (Ram and Sivaswamy [132])	0.72	-	0.87
Soft exudates	-	-	-
<u>Task: detect images with diabetic lesion type</u>			
Haemorrhages (Hatanaka et al. [72])	0.80	0.80	-
Microaneurysms (Fleming et al. [45])	0.85	0.83	-
Hard exudates (Sánchez et al. [142])	1.00	0.90	-
Soft exudates (Zhang and Chutatape [197, 199])	0.88	0.84	-

On pixel level, the baseline algorithm at first glance seems to achieve moderate results in hard exudate detection compared to the state-of-the-art method [132]. The positive

predictive value (PPV), however, indicates that the specificity of the state-of-the-art method would be much better than the baseline algorithm's specificity if it were reported. The baseline algorithm also shows worse performance than the state-of-the-art microaneurysm detection method [20] which is emphasised by the fact that the reported specificity is misleading. In the evaluation, the method used only fraction of the true negative image pixels which produced worse specificity for the method than it actual is. Hence, the performance of the baseline algorithm fall far behind in the pixel level performance compared to the state-of-the-art methods. Since most of the pixels in eye fundus images are not affected by diabetic retinopathy and to produce any meaningful segmentation the pixel level specificity should be close to one, it can be concluded that on pixel level the photometric cue is not good enough to detect diabetic retinopathy, and requires support from other features.

Apart from microaneurysm detection, the baseline algorithm produced slightly higher sensitivity and specificity values on image level than on pixel level. The interpretation of sensitivity and specificity, however, depend on how the test populations are balanced, and since the proportion of the lesioned images in the database is much better balanced than the proportion of the lesioned pixels in the database images, the baseline algorithm actually performed clearly better on image level. Against this background, the baseline algorithm managed to detect the lesioned images relatively well without introducing considerable number of false detections which is surprising considering the pixel performances. This is an important piece of information from the screening point of view. On the other hand, all the state-of-the-art methods [72, 45, 142, 197, 199] in Table 7.3 performed better than the baseline algorithm. Hence, the colour cue alone is not perhaps good enough to detect diabetic retinopathy on image level, but it should be definitely included as one of the features when designing automatic image analysis methods for screening diabetic retinopathy.

In the optic disc localisation in Chapter 6, the colour characteristics of an optic disc was captured using PCA to find the vector basis for the colour data extracted from several optic disc regions. The colour decorrelated templates were extracted from images projected to the space spanned by the resulting eigenvectors (i.e., the colour decorrelated template space). In the localisation stage, the input image was projected to the colour decorrelated template space, where the optic disc was located by using colour decorrelated templates and customised similarity metric in channel-wise template matching. The optic disc localisation also supported the discussion that illuminance correction is more useful when exploiting spatial dependency.

One limitation of this study was photometric information in eye fundus images. The dynamic range of the standard eye fundus cameras was not sufficient to record all the required information which appeared as overexposed and underexposed pixels in the images. Thus, the results do not represent the full potential of the photometric cue. Moreover, the photometric differences between eye fundus cameras restrict the use of photometric information in the lesion detection unless colour calibration is used. Therefore, the aim of future work is to improve the current image capturing process and support the development of new applied imaging technologies, such as spectral imaging of the eye fundus [37]. The improvements for current image capturing comprises colour calibration, dynamic range stretching (e.g. improving the imaging system or using high dynamic range imaging (HDR)), and automation of image capturing steps (e.g. focusing [108]).

- [1] Application protocol for electronic data exchange in healthcare environments version 2.5.1. ANSI Standard. Available online at: <http://www.hl7.org>.
- [2] Collection of multispectral images of the fundus (CMIF). Electronic material (Online). Available online at: <http://www.cs.bham.ac.uk/research/projects/fundus-multispectral/> [referred 16.12.2009].
- [3] Diabetic retinopathy database and evaluation protocol (DIARETDB0). Electronic material (Online). Available online at: <http://www.it.lut.fi/project/imageret/files/diaretdb0> [referred 19.2.2010].
- [4] Diabetic retinopathy database and evaluation protocol (DIARETDB1). Electronic material (Online). Available online at: http://www.it.lut.fi/project/imageret/files/diaretdb_v02_01/ [referred 18.12.2009].
- [5] Digital retinal images for vessel extraction (DRIVE). Electronic material (Online). Available online at: <http://www.isi.uu.nl/Research/Databases/DRIVE/> [referred 16.12.2009].
- [6] Image annotation tool (IMGANNOOTOOL). Electronic material (Online). Available online at: <http://www.it.lut.fi/project/imgannotool> [referred 18.12.2009].
- [7] Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology (MESSIDOR). Electronic material (Online). Available online at: <http://messidor.crihan.fr> [referred 16.12.2009].
- [8] Retinopathy online challenge (ROC). Electronic material (Online). Available online at: <http://roc.healthcare.uiowa.edu/> [referred 16.12.2009].
- [9] Structured analysis of the retina (STARE). Electronic material (Online). Available online at: <http://www.ces.clemson.edu/~ahoover/stare/> [referred 16.12.2009].
- [10] REVIEW: Retinal vessel image set for estimation of widths (REVIEW). Electronic material (Online). Available online at: <http://reviewdb.lincoln.ac.uk/> [referred 16.12.2009].
- [11] ABRÁMOFF, M. D., NIEMEIJER, M., SUTTORP-SCHULTEN, M. S. A., VIEGEVER, M. A., RUSSEL, S. R., AND VAN GINNEKEN, B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a

- large population of patients with diabetes. *Diabetes care* 31, 2 (February 2008), 193 – 197.
- [12] AGURTO, C., MURRAY, V., BARRIGA, E., SERGIO MURILLO, M. P., DAVIS, H., RUSSELL, S., ABRMOFF, M., AND SOLIZ, P. Multiscale am-fm methods for diabetic retinopathy lesion detection. *IEEE Transactions on Medical Imaging* 29, 2 (2010), 502 – 512.
- [13] AL-DIRI, B., HUNTER, A., STEEL, D., HABIB, M., HUDAIB, T., AND BERRY, S. REVIEW - a reference data set for retinal vessel profiles. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2262–2265.
- [14] ALDINGTON, S. J., KOHNER, E. M., MEUER, S., KLEIN, R., AND SJØOLIE, A. K. Methodology for retinal photography and assessment of diabetic retinopathy: the european community funded concerted action programme (EURODIAB) insulin-dependent diabetes mellitus IDDM complications study. *Diabetologia* 38, 4 (1995), 437–444.
- [15] ANITHA, J., SELVATHI, D., AND HEMANTH, D. J. Neural computing based abnormality detection in retinal optical images. In *Proceedings of the International Advanced Computing Conference (IACC 2009)* (2009), pp. 630 – 635.
- [16] BAILLIERE, E., BENGIO, S., BIMBOT, F., HAMOUZ, M., KITTLER, J., MARIETHOZ, J., MATAS, J., MESSER, K., POPOVICI, V., POREE, F., RUIZ, B., AND THIRAN, J. The BANCA database and evaluation protocol. In *Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA2003)* (2003), pp. 625–638.
- [17] BARRIGA, E. S., MURRAY, V., AGURTO, C., PATTICHIS, M., BAUMAN, W., ZAMORA, G., AND SOLIZ, P. Automatic system for diabetic retinopathy screening based on am-fm, partial least squares, and support vector machines. In *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2010), pp. 1349 – 1352.
- [18] BAUDOIN, C. E., LAY, B. J., AND KLEIN, J. C. Automatic detection of microaneurysms in diabetic fluorescein angiography. *Revue D'Épidémiologie et de Santé Publique* 32 (1984), 254 – 261.
- [19] BAUDOIN, C. E., MANESCHI, F., QUENTEL, G., SOUBRANE, G., HAYES, T., JONES, G., COSCAS, G., AND KOHNER, E. M. Quantitative evaluation of fluorescein angiograms: microaneurysm counts. 8 – 13.
- [20] BHALERAO, A., PATANAİK, A., ANAND, S., AND SARAVANAN, P. Robust detection of microaneurysms for sight threatening retinopathy screening. In *Proceedings of the Sixth Indian Conference on Computer Vision, Graphics and Image Processing (ICGVIP2008)* (2008), pp. 520 – 527.
- [21] BIANCARDI, A. M., AND REEVES, A. P. Tesd: a novel ground truth estimation method. In *Proceedings of the SPIE Medical Imaging 2009: Computer-Aided Diagnosis* (2009), vol. 7260.

- [22] BRADLEY, A. E. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 7 (1997), 1145 – 1159.
- [23] CARINCI, F., STRACCI, F., AND STORMS, F. Diabetes. In *Major and Chronic Diseases, Report 2007*. European Union, Directorate-General for Health and Consumers, 2007. ISBN 92-79-08896-4.
- [24] COOTES, T., TAYLOR, C., COOPER, D., , AND GRAHAM, J. Active shape models – their training and application. *Computer Vision and Image Understanding* 61, 1 (1995), 38–59.
- [25] CREE, M., OLSON, J., MCHARDY, K., FORRESTER, J., AND SHARP, P. Automated microaneurysm detection. In *Proceedings of the International Conference on Image Processing (ICIP1996)* (1996), pp. 699 – 702.
- [26] DAVID, J., , KRISHNAN, R., AND KUMAR, A. S. Neural network based retinal image analysis. In *Proceedings of the International Congress on Image and Signal Processing (CISP2008)* (2008), vol. 2, pp. 49 – 53.
- [27] DE LECLAIRAGE COMMISSION (CIE), C. I. Cie colorimetry part 4: 1976 L*a*b* colour space. Joint ISO/CIE Standard. ISO 11664-4:2008(E)/CIE S 014-4/E:2007.
- [28] DIABETES (ONLINE). Current Care Guideline. Working group set up by the Finnish Medical Society Duodecim and the Finnish Respiratory Society. Helsinki: Finnish Medical Society Duodecim, 2009 [referred 19.1.2010]. Available online at: www.kaypahoito.fi.
- [29] DIABETIC RETINOPATHY (ONLINE). Current Care Guideline. Working group set up by the Finnish Medical Society Duodecim and the Finnish Respiratory Society. Helsinki: Finnish Medical Society Duodecim, 2006 [referred 19.1.2010]. Available online at: www.kaypahoito.fi.
- [30] DODD, L. E., AND PEPE, M. S. Partial auc estimation and regression. *Biometrics* 59 (2003), 614 – 623.
- [31] DUPAS, B., WALTER, T., ERGINAY, A., ORDONEZ, R., DEB-JOARDAR, N., GAIN, P., KLEIN, J.-C., AND MASSIN, P. Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a computer-assisted diagnostic system for grading diabetic retinopathy. *Diabetes & Metabolism* 36 (2010), 213 – 220.
- [32] EGE, B. M., HEJLESEN, O. K., LARSEN, O. V., MLLER, K., JENNINGS, B., KERR, D., AND CAVAN, D. A. Screening for diabetic retinopathy using computer based image analysis and statistical classification. *Computer Methods and Programs in Biomedicine* 62 (2000), 165 – 175.
- [33] ENRICO GRISAN, A. R. Segmentation of candidate dark lesions in fundus images based on local thresholding and pixel density. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2007)* (2007), pp. 6735 – 6738.

- [34] ESTABRIDIS, K., AND DE FIGUEIREDO, R. J. P. Automatic detection and diagnosis of diabetic retinopathy. In *Proceedings of the International Conference on Image Processing (ICIP2007)* (2007), vol. 2, pp. 445 – 448.
- [35] ESWARAN, C., REZA, A. W., , AND HATI, S. Extraction of the contours of optic disc and exudates based on marker- controlled watershed segmentation. In *Proceedings of the International Conference on Computer Science and Information Technology (ICCSIT2008)* (2008), pp. 719 – 723.
- [36] EVERINGHAM, M., AND ZISSERMAN, A. The pascal visual object classes challenge 2006 (VOC2006) results. Workshop in the European Conference on Computer Vision (ECCV2006), May. Graz, Austria.
- [37] FÄLT, P., HILTUNEN, J., HAUTA-KASARI, M., SORRI, I., KALESNYKIENE, V., AND UUSITALO, H. Extending diabetic retinopathy imaging from color to spectra. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA2009)* (Oslo, Norway, June 2009), vol. 5575/2009 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 149 – 158.
- [38] FAUST, O., U., R. A., NG, E. Y. K., NG, K.-H., AND SURI, J. S. Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review. *Journal of Medical Systems* (2010), 1 – 13.
- [39] FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters* 27, 8 (June 2006), 861–874.
- [40] FEI-FEI, L., FERGUS, R., AND PERONA, P. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 4 (2006), 594–611.
- [41] FIGUEIREDO, M., AND JAIN, A. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 3 (Mar 2002), 381–396.
- [42] FINLAYSON, G. D., SCHIELE, B., AND CROWLEY, J. L. Comprehensive colour image normalization. In *Proceedings of the European Conference on Computer Vision (ECCV1998)* (Freiburg, Germany, June 1998), vol. 1, pp. 475–490.
- [43] FLEMING, A. D., GOATMAN, K. A., PHILIP, S., WILLIAMS, G. J., PRESCOTT, G. J., SCOTLAND, G. S., MCNAMEE, P., LEESE, G. P., WYKES, W. N., SHARP, P. F., AND OLSON, J. A. The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy. *British Journal of Ophthalmology* 94 (2009), 706 – 711.
- [44] FLEMING, A. D., GOATMAN, K. A., WILLIAMS, G. J., PHILIP, S., SHARP, P. F., AND OLSON, J. A. Automated detection of blot haemorrhages as a sign of referable diabetic retinopathy. In *Proceedings of the Medical Image Understanding and Analysis (MIUA2008)* (2008), pp. 235 – 239.
- [45] FLEMING, A. D., PHILIP, S., GOATMAN, K. A., OLSON, J. A., AND SHARP, P. F. Automated microaneurysm detection using local contrast normalization and

- local vessel detection. *IEEE Transactions On Medical Imaging* 25, 9 (2006), 1223 – 1232.
- [46] FLEMING, A. D., PHILIP, S., GOATMAN, K. A., WILLIAMS, G. J., OLSON, J. A., AND SHARP, P. F. Automated detection of exudates for diabetic retinopathy screening. *Physics In Medicine And Biology* 52 (2007).
- [47] FORACCHIA, M., GRISAN, E., AND RUGGERI, A. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transactions On Medical Imaging* 23, 10 (October 2004), 1189–1195.
- [48] FORACCHIA, M., GRISAN, E., AND RUGGERI, A. Luminosity and contrast normalization in retinal images. *Medical Image Analysis* 9, 3 (June 2005), 179–190.
- [49] FORRESTER, J. V., DICK, A. D., AND AND WILLIAM R. LEE, P. G. M. *The Eye, Basis sciences in practice*. Saunders Ltd, 2nd edition, 2001.
- [50] FRAME, A. J., UNDRILL, P. E., CREE, M. J., OLSON, J. A., MCHARDY, K. C., SHARP, P. F., AND FORRESTER, J. V. A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic uorescein angiograms. *Computers in Biology and Medicine* 28 (1998), 225 – 238.
- [51] FUMERA, G., AND ROLI, F. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 6 (June 2005), 942 – 956.
- [52] GARCÍA, M., HORNERO, R., SÁNCHEZ, C. I., LÓPEZ, M. I., AND DÍEZ, A. Feature extraction and selection for the automatic detection of hard exudates in retinal images. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2007)* (2007), pp. 4969 – 4972.
- [53] GARCÍA, M., SÁNCHEZ, C. I., LÓPEZ, M. I., DÍEZ, A., AND HORNERO, R. Automatic detection of red lesions in retinal images using a multilayer perceptron neural network. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2008)* (2008), pp. 5425 – 5428.
- [54] GARCÍA, M., SÁNCHEZ, C. I., POZA, J., LÓPEZ, M. I., AND HORNERO, R. Detection of hard exudates in retinal images using a radial basis function classifier. *Annals of Biomedical Engineering* 37, 7 (2009), 1448 – 1463.
- [55] GARDNER, G. G., KEATING, D., WILLIAMSON, T. H., AND ELLIOTT, A. T. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British Journal of Ophthalmology* 80 (1996), 940 – 944.
- [56] GILL, C., SABIN, L., AND SCHMID, C. Why clinicians are natural bayesians. *British Medical Journal* 330, 7 (2005), 1080–1083.
- [57] GOATMAN, K. A., CREE, M. J., OLSON, J. A., FORRESTER, J. V., AND SHARP, P. F. Automated measurement of microaneurysm turnover. *Investigative Ophthalmology and Visual Science* 44, 12 (2003), 5335 – 5341.

- [58] GOATMAN, K. A., WHITWAM, A. D., MANIVANNAN, A., OLSON, J. A., AND SHARP, P. F. Colour normalisation of retinal images. In *Proceedings of the Medical Image Understanding and Analysis (MIUA2003)* (2003).
- [59] GOH, J., TANG, L., SALEH, G., TURK AND YU FU, L. A., AND BROWNE, A. Filtering normal retinal images for diabetic retinopathy screening using multiple classifiers. In *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB2009)* (2009), pp. 1 – 4.
- [60] GOH, K. G., HSU, W., LEE, M. L., AND WANG, H. Adris: an automatic diabetic retinal image screening system. In *Medical Data Mining and Knowledge Discovery*, vol. 60 of *Studies in Fuzziness and Soft Computing*. 2001.
- [61] GOLDBAUM, M., BOYD, S. M. A. T. S. C. J., HUNTER, E., AND JAIN, R. Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images. In *Proceedings of the International Conference on Image Processing (ICIP1996)* (1996), vol. 3, pp. 695 – 698.
- [62] GOLDBAUM, M. H., KATZ, N. P., CHAUDHURI, S., AND NELSON, M. Image understanding for automated retinal diagnosis. In *Proceedings of the Annual Symposium on Computer Application in Medical Care* (1989), vol. 8, pp. 756 – 760.
- [63] GOLDBAUM, M. H., KATZ, N. P., NELSON, M. R., AND HAFF, L. R. The discrimination of similarly colored objects in computer images of the ocular fundus. *Investigative Ophthalmology & Visual Science* 31 (1990).
- [64] GRISAN, E., GIANI, A., CESERACCIU, E., AND RUGGERI, A. Model-based illumination correction in retinal images. In *Proceedings of the IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI2006)* (April 2006), pp. 984–987.
- [65] GROSVENOR, T. *Primary care optometry*. Elsevier Health Sciences, 2006.
- [66] GROTHOR, P., MICHEALS, R., AND PHILLIPS, P. Face recognition vendor test 2002 performance metrics. In *Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA2003)* (2003), pp. 937–945.
- [67] GRUNDLAND, M., AND DODGSON, N. A. Color histogram specification by histogram warping. In *Proceedings of the SPIE on Color Imaging X: Processing, Hardcopy, and Applications* (2005), vol. 5667, pp. 837–840.
- [68] GYUTON, A. C., AND HALL, J. E. *Textbook of medical physiology, 9th edition*. W.B Saunders Company, 1996.
- [69] HAND, D. J. Evaluating diagnostic tests: The area under the roc curve and the balance of errors. *Statistics in Medicine* 29 (2010), 1502 – 1510.
- [70] HANLEY, J. A., AND MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* 143 (1982), 39 – 36.

- [71] HANSEN, A. B., HARTVIG, N. V., JENSEN, M. S., BORCH-JOHNSEN, K., LUND-ANDERSEN, H., AND LARSEN, M. Diabetic retinopathy screening using digital non-mydratic fundus photography and automated image analysis. *Acta Ophthalmologica Scandinavica* 82 (2004).
- [72] HATANAKA, Y., NAKAGAWA, T., HAYASHI, Y., HARA, T., AND FUJITA, H. Improvement of automated detection method of hemorrhages in fundus images. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008)* (2008), pp. 5429 – 5432.
- [73] HATANAKA, Y., NAKAGAWA, T., HAYASHI, Y., MIZUKUSA, Y., FUJITA, A., KAKOGAWA, M., AND TAKESHI HARA, K. K., AND FUJITA, H. Cad scheme to detect hemorrhages and exudates in ocular fundus images. In *Proceedings of the SPIE Medical Imaging 2007: Computer-Aided Diagnosis* (2007), vol. 6514.
- [74] HIPWELL, J. H., STRACHAN, F., OLSON, J. A., MCHARDY, K. C., SHARP, P. F., AND FORRESTER, J. V. Automated detection of microaneurysms in digital red-free photographs: a diabetic screening tool. *Diabetic medicine* 17 (2000), 588 – 594.
- [75] HOOVER, A., AND GOLDBAUM, M. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transactions on Medical Imaging*, 22, 8 (August 2003), 951–958.
- [76] HOOVER, A., KOUZNETSOVA, V., AND GOLDBAUM, M. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging* 19, 3 (March 2000), 203–210.
- [77] HSU, W., PALLAWALA, P. M. D. S., LEE, M. L., AND EONG, K.-G. A. The role of domain knowledge in the detection of retinal hard exudates. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2001)* (2001), vol. 2, pp. 246 – 251.
- [78] HUANG, D., KAISER, P. K., LOWDER, C. Y., AND TRABOULSI, E. I. *Retinal imaging*. Mosby Elsevier, 2006.
- [79] HUTCHINSON, A., MCINTOSH, A., PETERS, J., O’KEEFFE, C., KHUNTI, K., BAKER, R., AND BOOTH, A. Effectiveness of screening and monitoring tests for diabetic retinopathy – a systematic review. *Diabetic medicine* 17, 7 (July 2000), 495–506.
- [80] KAHAI, P., NAMUDURI, K. R., AND THOMPSON, H. Decision support for automated screening of diabetic retinopathy. In *Proceedings of the International Asilomar Conference of the IEEE Signals, Systems and Computers (SSC2004)* (2004), vol. 2, pp. 1630 – 1634.
- [81] KAHAI, P., NAMUDURI, K. R., AND THOMPSON, H. A decision support framework for automated screening of diabetic retinopathy. *International Journal of Biomedical Imaging* (2006), 1 – 8.

- [82] KAMARAINEN, J.-K., HAMOUZ, M., KITTLER, J., PAALANEN, P., ILONEN, J., AND DROBCHENKO, A. Object localisation using generative probability model for spatial constellation and local image features. In *In Proceedings of the International Conference on Computer Vision (ICCV2007), Workshop on Non-Rigid Registration and Tracking Through Learning* (October 2007), pp. 1–8.
- [83] KAMEL, M., BELKASSIM, S., MENDONCA, A. M., AND CAMPILHO, A. A neural network approach for the automatic detection of microaneurysms in retinal angiograms. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN2001)* (2001), vol. 4, pp. 2695 – 2699.
- [84] KANG, S. B., AND WEISS, R. Can we calibrate a camera using an image of a flat, textureless lambertian surface? In *Proceedings of the European Conference on Computer Vision (ECCV2000)* (2000), pp. 640–653.
- [85] KAUPPI, T., KALESNYKIENE, V., KAMARAINEN, J.-K., LENSU, L., SORRI, I., A. RANINEN, A., VOUTILAINEN, R., UUSITALO, H., KÄLVIÄINEN, H., AND PIETILÄ, J. The diaretdb1 diabetic retinopathy database and evaluation protocol. In *Proceedings of the British Machine Vision Conference (BMVC2007)* (Warwick, UK, 2007), vol. 1, pp. 252–261.
- [86] KAUPPI, T., KALESNYKIENE, V., KAMARAINEN, J.-K., LENSU, L., SORRI, I., RANINEN, A., VOUTILAINEN, R., PIETILÄ, J., KÄLVIÄINEN, H., AND UUSITALO, H. DIARETDB1 diabetic retinopathy database and evaluation protocol. In *Proceedings of the Medical Image Understanding and Analysis (MIUA2007)* (Aberystwyth, UK, 2007), pp. 61–65.
- [87] KAUPPI, T., KALESNYKIENE, V., KAMARAINEN, J.-K., LENSU, L., SORRI, I., UUSITALO, H., KÄLVIÄINEN, H., AND PIETILÄ, J. DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms, 2006. Published online at: <http://www2.it.lut.fi/project/imageret/diaretdb0/>.
- [88] KAUPPI, T., AND KÄLVIÄINEN, H. Simple and robust optic disc localisation using colour decorrelated templates. In *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS2009)* (Juan-les-Pins, France, 2008), vol. 5259, pp. 719 – 729.
- [89] KAUPPI, T., KAMARAINEN, J.-K., LENSU, L., KALESNYKIENE, V., SORRI, I., KÄLVIÄINEN, H., UUSITALO, H., AND PIETILÄ, J. Fusion of multiple expert annotations and overall score selection for medical image diagnosis. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA2009)* (Oslo, Norway, 2009), pp. 760–769.
- [90] KAUPPI, T., LENSU, L., KÄMÄRÄINEN, J.-K., FÄLT, P., HAUTA-KASARI, M., HILTUNEN, J., KALESNYKIENE, V., SORRI, I., UUSITALO, H., KÄLVIÄINEN, H., AND PIETILÄ, J. The role of model-based illumination correction in processing colour eye fundus images. In *Proceedings of the International Union for Physical and Engineering Sciences in Medicine (World Congress on Medical Physics and Biomedical Engineering (WC2009))* (Munich, Germany, 2009).

- [91] KIM, S. J., AND POLLEFEYS, M. Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 4 (April 2008), 563–576.
- [92] KITTLER, J., HATEF, M., DUIN, R. P. W., AND MATAS, J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 3 (March 1998), 226 – 239.
- [93] KRISTINSSON, J. K., HAUKSDOTTIR, H., STEFANSSON, E., JONASSON, F., AND GISLASON, I. Active prevention in diabetic eye disease. *Acta ophthalmologica Scandinavica* 75, 3 (June 1997), 249–54.
- [94] KUMARI, V. V., SURIYAHARAYANANM, N., AND SARANYA, C. T. Feature extraction for early detection of diabetic retinopathy. In *Proceedings of the International Conference on Recent Trends in Information, Telecommunication and Computing (ITC2010)* (2010), pp. 359 – 361.
- [95] LALONDE, M., BEAULIEU, M., AND GAGNON, L. Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. *IEEE Transactions On Medical Imaging* 20, 11 (November 2001), 1193–1200.
- [96] LARSEN, N., GODT, J., GRUNKIN, M., LUND-ANDERSEN, H., AND LARSEN, M. Automated detection of diabetic retinopathy in a fundus photographic screening population. *Investigative Ophthalmology & Visual Science* 44, 2 (2003).
- [97] LASKO, T. A., BHAGWAT, J. G., ZOU, K. H., AND OHNO-MACHADO, L. The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics* 38 (2005), 404–415.
- [98] LI, H., AND CHUTATAPE, O. Automatic location of optic disk in retinal images. In *Proceedings of the International Conference on Image Processing (ICIP2001)* (Thessaloniki, Greece, 2001), vol. 2, pp. 837–840.
- [99] LI, H., AND CHUTATAPE, O. A model-based approach for automated feature extraction in fundus images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV2003)* (2003), vol. 1, pp. 394 – 399.
- [100] LI, H., AND CHUTATAPE, O. Automated feature extraction in color retinal images by a model based approach. *IEEE Transactions On Biomedical Engineering* 51, 2 (2004), 246 – 254.
- [101] LI, X., ALDRIDGE, B., REES, J., AND FISHER, R. Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation. In *Proceedings of the Medical Image Understanding and Analysis* (2010), vol. 1, pp. 101 – 106.
- [102] LUO, G., CHUTATAPE, O., LEI, H., AND KRISHNAN, S. M. Abnormality detection in automated mass screening system of diabetic retinopathy. In *Proceedings of the IEEE Symposium on Computer-Based Medical Systems (CBMS2001)* (2001), pp. 132 – 137.

- [103] MALACARA, D. *Color Vision and Colorimetry: Theory and Applications*. SPIE – International Society for Optical Engineering, 2002.
- [104] MARGOLIS, R., AND KAISER, K. Diagnostic modalities in diabetic retinopathy. In *Diabetic retinopathy*, E. Duh, Ed. Humana Press, 2008, pp. 109–133.
- [105] MARTINKAUPPI, B. *Face colour under varying illumination - analysis and applications*. PhD thesis, University of Oulu, 2002.
- [106] MENDONCA, A. M., CAMPILHO, A. J., AND NUNES, J. M. Automatic segmentation of microaneurysms in retinal angiograms of diabetic patients. In *Proceedings of the International Conference on Image Analysis and Processing (ICIAP1999)* (1999), pp. 728 – 733.
- [107] MEYERLE, C. B., CHEW, E. Y., AND FERRIS III, F. L. Nonproliferative diabetic retinopathy. In *Diabetic retinopathy*, E. Duh, Ed. Humana Press, 2008, pp. 109–133.
- [108] MOSCARITOLO, M., JAMPPEL, H., KNEZEVIČ, F., AND ZEIMER, R. An image based auto-focusing algorithm for digital fundus photography. *IEEE Transaction on Medical Imaging* 28, 11 (November 2009), 1703–1707.
- [109] NARASIMHA-IYER, H., CAN, A., ROYSAM, B., STEWART, C. V., TANENBAUM, H. L., MAJEROVIC, A., AND SINGH, H. Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy. *IEEE Transactions on Biomedical Engineering* 53, 6 (June 2006), 1084 – 1098.
- [110] NATIONAL EYE INSTITUTE, NATIONAL INSTITUTES OF HEALTH. Available online at: <http://www.nei.nih.gov/>. [referred 19.1.2010].
- [111] NAYAK, J., BHAT, P. S., U, R. A., LIM, C. M., AND KAGATHI, M. Automated identification of diabetic retinopathy stages using digital fundus images. *Journal of Medical Systems* 32 (2008), 107 – 115.
- [112] NIEMEIJER, M., ABRÁMOFF, M. D., AND VAN GINNEKEN, B. Information fusion for diabetic retinopathy cad in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 28, 5 (2009), 775 – 785.
- [113] NIEMEIJER, M., ABRMOFF, M. D., AND VAN GINNEKEN, B. Segmentation of the optic disc, macula and vascular arch in fundus photographs. *IEEE Transactions On Medical Imaging* 26, 1 (January 2007), 116–127.
- [114] NIEMEIJER, M., STAAL, J., VAN GINNEKEN, B., LOOG, M., AND ABRÁMOFF, M. Comparative study of retinal vessel segmentation a new publicly available database. In *Proceedings of the SPIE Medical Imaging* (2004), pp. 648–656.
- [115] NIEMEIJER, M., VAN GINNEKEN, B., CREE, M. J., MIZUTANI, A., QUELLEC, G., SANCHEZ, C. I., BOB ZHANG, R. H., LAMARD, M., MURAMATSU, C., WU, X., CAZUGUEL, G., YOU, J., MAYO, A., LI, Q., HATANAKA, Y., COCHENER, B., ROUX, C., , GARCIA, F. K. M., , FUJITA, H., AND ABRÁMOFF, M. D. Retinopathy online challenge: Automatic of microaneurysms in digital photographs. *IEEE Transactions on Medical Imaging* 29, 1 (January 2010), 185–195.

- [116] NIEMEIJER, M., VAN GINNEKEN, B., RUSSELL, S. R., AND ABRÁMOFF, M. S. A. S.-S. M. D. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative Ophthalmology and Visual Science* 48, 5 (2007), 2260–2267.
- [117] NIEMEIJER, M., VAN GINNEKEN, B., STAAL, J., SUTTORP-SCHULTEN, M. S. A., AND ABRÁMOFF, M. D. Automatic detection of red lesions in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 24, 5 (May 2005), 584 – 592.
- [118] NORONHA, K., NAYAK, J., AND BHAT, S. Enhancement of retinal fundus image to highlight the features for detection of abnormal eyes. In *Proceedings of the IEEE Region 10 Conference (TENCON2006)* (2006), pp. 1 – 4.
- [119] OSAREH, A. *Automated Identification of Diabetic Retinal Exudates and the Optic Disc*. PhD thesis, Department of Computer Science, University of Bristol, 2004.
- [120] OSAREH, A., MIRMEHDI, M., THOMAS, B., AND MARKHAM, R. Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks. In *Proceedings of the Medical Image Understanding and Analysis (MIUA2001)* (2001), pp. 49 – 52.
- [121] OSAREH, A., MIRMEHDI, M., THOMAS, B., AND MARKHAM, R. Classification and localisation of diabetic-related eye disease. In *Proceedings of the European Conference on Computer Vision (ECCV2002)* (May 2002), pp. 502–516.
- [122] OSAREH, A., SHADGAR, B., AND MARKHAM, R. A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. *IEEE Transactions On Information Technology In Biomedicine* 13, 4 (2009), 535 – 545.
- [123] OTANI, T., KISHI, S., AND MARUYAMA, Y. Patterns of diabetic macular edema with optical coherence tomography. *American journal of Ophthalmology* 127, 6 (1999), 88–693.
- [124] PAALANEN, P., KAMARAINEN, J.-K., ILONEN, J., AND KÄLVIÄINEN, H. Feature representation and discrimination based on Gaussian mixture model probability densities - practices and algorithms. *Pattern Recognition* 39, 7 (2006), 1346–1358.
- [125] PALLAWALA, P. M. D. S., HSU, W., LEE, M. L., AND GOH, S. S. Automated microaneurysm segmentation and detection using generalized eigenvectors. In *Proceedings of the Seventh IEEE Workshop on Applications of Computer Vision (WACV/MOTION2005)* (2005), vol. 1, pp. 322 – 327.
- [126] PARK, M., JIN, J. S., AND SUHUI. Locating the optic disc in retinal images. In *Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation (CGIV2006)* (Sydney, Australia, 2006), pp. 141–145.
- [127] PATTON, N., ASLAM, T. M., MACGILLIVRAY, T., DEARY, I. J., DHILLON, B., EIKELBOOM, R. H., YOGESAN, K., AND CONSTABLE, I. J. Retinal image analysis: Concepts, applications and potential. *Progress in Retinal and Eye Research* 25, 1 (January 2005), 99 –127.

- [128] PEPE, M. S. *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, 2003.
- [129] PHILLIPS, R., FORRESTER, J., AND SHARP, P. Automated detection and quantification of retinal exudates. *Graefe's Archive for Clinical and Experimental Ophthalmology* 231, 2 (1993), 90 – 94.
- [130] QUELLEC, G., LAMARD, M., JOSSELIN, P. M., CAZUGUEL, G., COCHENER, B., AND ROUX, C. Detection of lesions in retina photographs based on the wavelet transform. In *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2006)* (2006), pp. 2619 – 2621.
- [131] QUELLEC, G., LAMARD, M., JOSSELIN, P. M., CAZUGUEL, G., COCHENER, B., AND ROUX, C. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Transactions On Medical Imaging* 27, 9 (September 2008), 1230 – 1241.
- [132] RAM, K., AND SIVASWAMY, J. Multi-space clustering for segmentation of exudates in retinal color photographs. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2009)* (2009), pp. 1437 – 1440.
- [133] RAVISHANKAR, S., JAIN, A., AND MITTAL, A. Automated feature extraction for early detection of diabetic retinopathy in fundus images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2009)* (2009), pp. 210 – 217.
- [134] REINHARD, E., ASHIKHMIN, M., GOOCH, B., AND SHIRLEY, P. Color transfer between images. *IEEE Computer Graphics and Applications* 21, 5 (2001), 34–41.
- [135] REISS, M. Notes on the cos4 law of illumination. *Journal Of The Optical Society Of America* 38, 11 (November 1948), 980–986.
- [136] REZA, A. W., AND ESWARAN, C. A decision support system for automatic screening of non-proliferative diabetic retinopathy. *Journal of Medical Systems* 4 (2009), 1 – 8.
- [137] RUDERMAN, D. L. Statistics of cone responses to natural images: implications for visual coding. *Journal Of The Optical Society Of America* 15, 8 (1998), 2036 – 2045.
- [138] RUSS, J. C. *The image processing handbook, fourth edition*. CRC Press, 2002.
- [139] RUSSELL, B., TORRALBA, A., MURPHY, K., AND FREEMAN, W. LabelMe: A database and Web-based tool for image annotation. *International Journal of Computer Vision* 77, 1-3 (2008), 157–173.
- [140] SAGAR, A. V., BALASUBRAMANIAM, S., AND CHANDRASEKARAN, V. A novel integrated approach using dynamic thresholding and edge detection (idted) for automatic detection of exudates in digital fundus retinal images. In *Proceedings of the International Conference on Computing: Theory and Applications (ICCTA2007)* (2007), pp. 705 – 710.

- [141] SANCHEZ, C., HORNERO, R., LOPEZ, M., AND POZA, J. Retinal image analysis to detect and quantify lesions associated with diabetic retinopathy. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2004)* (2004), vol. 1, pp. 1624 – 1627.
- [142] SÁNCHEZ, C. I., GARCÍA, M., MAYO, A., LÓPEZ, M. I., AND HORNERO, R. Retinal image analysis based on mixture models to detect hard exudates. *Medical Image Analysis* 13 (2009), 650658.
- [143] SÁNCHEZ, C. I., MAYO, A., GARCÍA, M., LÓPEZ, M. I., AND HORNERO, R. Automatic image processing algorithm to detect hard exudates based on mixture models. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS2006)* (2006), pp. 4453 – 4456.
- [144] SÁNCHEZ, C. I., NIEMEIJER, M., SCHULTEN, M. S. A. S., ABRÁMOFF, M., AND VAN GINNEKEN, B. Improving hard exudate detection in retinal images through a combination of local and contextual information. In *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2010), pp. 5 – 8.
- [145] SAVOLAINEN, E., AND LEE, Q. Diabetic retinopathy - need and demand for photocoagulation and its cost-effectiveness: evaluation based services in the United Kingdom. *Diabetologia* 23, 2 (August 1982), 138–140.
- [146] SCHARCANSKI, D. W. J., AND MARINHO, D. R. A coarse-to-ne strategy for automatically detecting exudates in color eye fundus images. *Computerized Medical Imaging and Graphics* 34 (2010), 228 – 235.
- [147] SCHARSTEIN, D., AND SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision* 47, 1/2/3 (April–June 2002), 7–42.
- [148] SCOTT, D. *Multivariate Density Estimation: Theory , Practice and Visualization*. Wiley, 1992. New York.
- [149] SILBERMAN, N., AHRlich, K., FERGUS, R., AND SUBRAMANIAN, L. Case for automated detection of diabetic retinopathy. In *Proceedings of the AAAI Spring Symposium Series* (2010).
- [150] SINTHANAYOTHIN, C., BOYCE, J. F., COOK, H. L., AND WILLIAMSON, T. H. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology*, 83, 8 (August 1999), 902–910.
- [151] SINTHANAYOTHIN, C., BOYCE, J. F., WILLIAMSON, T. H., COOK, H. L., MENSAH, E., LAL, S., AND USHER, D. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Medicine* 19 (2002), 105 – 112.
- [152] SINTHANAYOTHIN, C., KONGBUNKIAT, V., PHOOJARUENCHANACHAI, S., AND SINGALAVANIJA, A. Automated screening system for diabetic retinopathy. In *Proceedings of the International Symposium on Image and Signal Processing and Analysis (ISPA2003)* (2003), vol. 2, pp. 915 – 920.

- [153] SIVAK-CALLCOTT, J. A., ODAY, D. M., GASS, J. D. M., AND TSAI, J. C. Evidence-based recommendations for the diagnosis and treatment of neovascular glaucoma. *Ophthalmology* 108, 10 (October 2001), 1767 – 1776.
- [154] SMITH, A. R. Color gamut transform pairs. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques* (1978), pp. 12 – 19.
- [155] SOARES, J., LEANDRO, J., CESAR, R., JELINK, H., AND CREE, M. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging* 25, 9 (September 2006), 1214–1222.
- [156] SPENCER, T., OLSON, J. A., MCHARDY, K. C., SHARP, P. F., AND FORRESTER, J. V. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Computers And Biomedical Research* 29 (1996), 284 – 302.
- [157] SPENCER, T., PHILLIPS, R. P., SHARP, P. F., AND FORRESTER, J. V. Automated detection and quantification of microaneurysms in fluorescein angiograms. *Graefe's Archive Ior Clinical and Experimental Ophthalmology* 230 (1992), 36 – 41.
- [158] STAAL, J., ABRÁMOFF, M., NIEMEIJER, M., VIERGEVER, M., AND VAN GINNEKEN, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging* 23, 4 (April 2004), 501–509.
- [159] STEFÁNSSON, E., BEK, T., PORTA, M., LARSEN, N., KRISTINSSON, J. K., AND AGARDH, E. Screening and prevention of diabetic blindness. *Acta Ophthalmologica Scandinavica* 78, 4 (December 2000), 374 – 385.
- [160] STYLES, I., CALCAGNI, A., CLARIDGE, E., ORIHUELA-ESPINA, F., AND GIBSON, J. Quantitative analysis of multi-spectral fundus images. *Medical Image Analysis* 10, 4 (August 2006), 578–597.
- [161] TAX, D. M. J., VAN BREUKELEN, M., DUIN, R. P. W., AND KITTLER, J. Combining multiple classifiers by averaging or by multiplying? *Pattern Recognition* 33, 9 (2000), 1475 – 1485.
- [162] TAYLOR, R. *Handbook of Retinal Screening in Diabetes*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO10 8SQ, England, 2006.
- [163] THACKER, N., CLARK, A., BARRON, J., BEVERIDGE, J., COURTNEY, P., CRUM, W., RAMESH, V., AND CLARK, C. Performance characterization in computer vision: A guide to best practices. *Computer Vision and Image Understanding* 109 (2008), 305–334.
- [164] THEODORIDIS, S., AND KOUTROUMBAS, K. *Pattern Recognition*. Academic Press, 1999.
- [165] TOBIN, K. W., CHAUM, E., GOVINDASAMY, V. P., AND KARNOWSKI, T. P. Detection of anatomic structures in human retinal imagery. *IEEE Transactions On Medical Imaging* 26, 12 (December 2007), 1729–1739.

- [166] TOMAŽEVIČ, D., LIKAR, B., AND PERNUŠ, F. Comparative evaluation of retrospective shading correction methods. *Journal of Microscopy* 208 (December 2000), 212–223.
- [167] TRUCCO, E., AND KAMAT, P. J. Locating the optic disk in retinal images via plausible detection and constraint satisfaction. In *Proceedings of the International Conference on Image Processing (ICIP2004)* (Singapore, 2004), pp. 837–840.
- [168] TRUITT, P. W., SOLIZ, P., FARNATH, D., AND NEMETH, S. Utility of color information for segmentation of digital retinal images: Neural network-based approach. In *Proceedings of the SPIE Conference on Image Processing* (February 1998), pp. 1470–1481.
- [169] TSIN, Y., RAMESH, V., AND KANADE, T. Statistical calibration of CCD imaging process. In *Proceedings of the International Conference on Computer Vision (ICCV2001)* (2001), vol. 1, pp. 480–487.
- [170] TURK, M. A., AND PENTLAND, A. P. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3, 1 (1991), 71–86.
- [171] TURK, M. A., AND PENTLAND, A. P. Face recognition using eigenfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR1991)* (June 1991), pp. 586–591.
- [172] U., R. A., YUN, W. L., NG, E. Y. K., YU, W., AND SURI, J. S. Imaging systems of human eye: A review. *Journal of Medical Systems* 32, 4 (2008), 301–315.
- [173] USHER, D., DUMSKYJ, M., HIMAGA, M., WILLIAMSON, T. H., NUSSEY, S., AND BOYCE, J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. *Diabetic Medicine* 21 (2003), 84–90.
- [174] VON WENDT, G. *Screening for diabetic retinopathy: Aspects of photographic methods*. PhD thesis, Karolinska Institutet, November 2005.
- [175] VON WENDT, G., HEIKKILÄ, K., AND SUMMANEN, P. Detection of retinal neovascularizations using 45° and 60° photographic fields with varying 45° fields simulated on a 60° photograph. *Acta Ophthalmologica Scandinavica* 80, 4 (2002), 372–378.
- [176] VON WENDT, G., SUMMANEN, P., HALLNÄS, K., ALGVERE, P., HEIKKILÄ, K., AND SEREGARD, S. Detection of diabetic retinopathy: a comparison between red-free digital images and colour transparencies. *Graefe's Archive for Clinical and Experimental Ophthalmology* 243, 5 (May 2004), 427–432.
- [177] WADE, N. J. Image, eye, and retina (invited review). *Journal of the Optical Society of America. A, Optics, image science, and vision* 24, 5 (2007), 1229–1249.
- [178] WALTER, T., AND KLEIN, J.-C. Automatic detection of microaneurysms in color fundus images of the human retina by means of the bounding box closing. In *Medical Data Analysis* (2002), pp. 210–220.
- [179] WALTER, T., AND KLEIN, J.-C. A computational approach to diagnosis of diabetic retinopathy. In *Proceedings of the 6th Conference on Systemics, Cybernetics and Informatics (SCI2002)* (2002), pp. 521–526.

- [180] WALTER, T., KLEIN, J.-C., MASSIN, P., AND ERGINAY, A. A contribution of image processing to the diagnosis of diabetic retinopathy – detection of exudates in color fundus images of the human retina. *IEEE Transactions On Medical Imaging* 21, 10 (2002), 1236 – 1243.
- [181] WALTER, T., MASSIN, P., ERGINAY, A., A, R. O., JEULIN, C., AND KLEIN, J.-C. Automatic detection of microaneurysms in color fundus images. *Medical Image Analysis* 11 (2007), 555 – 566.
- [182] WANG, H., HSU, W., GOH, K. G., AND LEE, M. L. An effective approach to detect lesions in color retinal images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2000)* (2000), vol. 2, pp. 181 – 186.
- [183] WARD, N., TOMLINSON, S., AND TAYLOR, C. Image analysis of fundus photographs. the detection and measurement of exudates associated with diabetic retinopathy. *Ophthalmology* 96, 1 (1989), 80 – 86.
- [184] WARFIELD, S., ZOU, K., AND WELLS, W. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging* 23, 7 (2004), 903 – 921.
- [185] WIKIPEDIA. Eye — wikipedia, the free encyclopedia. Electronic material (Online).
- [186] WILKINSON, C. P., FERRIS, F. L., KLEIN, R. E., LEE, P. P., AGARDH, C. D., DAVIS, M., DILLS, D., KAMPIK, A., PARARAJASEGARAM, R., AND VERDAGUER, J. T. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 110, 9 (September 2003), 1677 – 1682.
- [187] WILSON, J. M. G., AND JUNGNER, G. Principles and practice of screening for disease. Tech. rep., World Health Organization Geneva, 1968.
- [188] WINDER, R., MORROW, P., MCRITCHIE, I., BAILIE, J., AND HART, P. Algorithms for digital image processing in diabetic retinopathy. *Computerized Medical Imaging and Graphics* 33, 8 (December 2009), 608–622.
- [189] WINELL, K., AND REUNANEN, A. Diabetes barometer 2005, 2006. ISBN 952-486-023-6.
- [190] WORLD HEALTH ORGANIZATION. Definition, diagnosis and classification of diabetes mellitus and its complications: Part 1: Diagnosis and classification of diabetes mellitus. Tech. rep., World Health Organization Noncommunicable Geneva, 1999. Report of a WHO Consultation.
- [191] WORLD HEALTH ORGANIZATION AND THE INTERNATIONAL DIABETES FEDERATION. Diabetes action now: An initiative of the world health organization and the international diabetes federation, 2004. ISBN 92 4 159151.
- [192] XIE, X., AND MIRMEHDI, M. Localising surface defects in random colour textures using multiscale texem analysis in image eigenchannels. In *Proceedings of the International Conference on Image Processing (ICIP2005)* (Genova, Italy, 2005), vol. 3, pp. 837–840.

- [193] XU, L., AND LUO, S. Support vector machine based method for identifying hard exudates in retinal images. In *Proceedings of the IEEE Youth Conference on Information, Computing and Telecommunication (YC-ICT2009)* (2009), pp. 138 – 141.
- [194] YAGMUR, F. D., KARLIK, B., AND OKATAN, A. Automatic recognition of retinopathy diseases by using wavelet based neural network. In *Proceedings of the International Conference on Applications of Digital Information and Web Technologies (ICADIWT2008)* (2008), pp. 454 – 457.
- [195] YANG, G., GAGNON, L., WANG, S., AND BOUCHE, M.-C. Algorithm for detecting micro-aneurysms in low-resolution color retinal images. In *Proceedings of Vision Interfaces (VI2001)* (2001), pp. 265 – 271.
- [196] YUN, W. L., ACHARYA, U. R., VENKATESH, Y., CHEE, C., MIN, L. C., AND NG, E. Identification of different stages of diabetic retinopathy using retinal optical image. *Information Sciences 178* (2008), 106 –121.
- [197] ZHANG, X., AND CHUTATAPE, O. Detection and classification of bright lesions in color fundus images. In *Proceedings of the International Conference on Image Processing (ICIP2004)* (2004), vol. 1, pp. 139 – 142.
- [198] ZHANG, X., AND CHUTATAPE, O. A svm approach for detection of hemorrhages in background diabetic retinopathy. In *Proceedings of International Joint Conference on Neural Networks* (2005), vol. 4, pp. 2435 – 2440.
- [199] ZHANG, X., AND CHUTATAPE, O. Top-down and bottom-up strategies in lesion detection of background diabetic retinopathy. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005)* (June 2005), vol. 2, pp. 422 – 428.
- [200] ZHENG, L., CHUTATAPE, O., AND SHANKAR, M. K. Automatic image analysis of fundus photograph. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS1997)* (1997), vol. 2, pp. 524 – 525.
- [201] ZOU, K. Receiver operating characteristic (ROC) literature research. Electronic material (Online). Available online at: <http://splweb.bwh.harvard.edu:8000/pages/pp1/zou/roc.html> [referred 15.4.2010].

Database characteristics

Table A.1: Train and test characteristics of the DIARETDB0 database, where HA = haemorrhages, MA = microaneurysms, HE = hard exudates, SE = soft exudates, NV = neovascularisation and NR = normal images (without diabetic lesions).

Set	No. train images	No. test images	Train sets (no. lesioned/normal images)						Test sets (no. lesioned/normal images)					
			HA	MA	HE	SE	NV	NR	HA	MA	HE	SE	NV	NR
1	10	120	6	7	4	3	2	3	74	99	67	38	18	19
2	20	110	12	15	11	8	4	5	68	91	60	33	16	17
3	30	100	19	23	15	10	6	7	61	83	56	31	14	15
4	40	90	26	32	17	14	8	8	54	74	54	27	12	14
5	50	80	32	39	25	16	10	10	48	67	46	25	10	12
6	60	70	36	48	30	20	12	12	44	58	41	21	8	10
7	70	60	39	53	34	25	14	16	41	53	37	16	6	6
8	80	50	47	61	39	27	17	17	33	45	32	14	3	5
9	90	40	51	69	45	28	19	19	29	37	26	13	1	3

A. Database characteristics

Table A.2: Train set characteristics of the DIARETDB1 database comprising expert annotations and the DIARETDB1 ground truth (result of expert annotation fusion), where HA = haemorrhages, where MA = mircoaneurysms, HE = hard exudates, SE = soft exudates, IRMA = intra retinal microvascular abnormalities, NV = neovascularisation and OD = optic disc.

Train set (28 images)							
No.	HA	MA	HE	SE	IRMA	NV	OD
Expert 1 (1 image without annotated lesions*)							
Annotated images	24	24	15	8	8	1	-
Annotated lesions	282	219	64	16	17	4	-
Annotated pixels	1657523	95127	1956913	150333	327715	124138	-
Expert 2 (3 images without annotated lesions*)							
Annotated images	22	20	18	13	3	4	26
Annotated lesions	233	263	84	37	5	16	26
Annotated pixels	1933697	2679277	2489881	290706	147791	962963	1075100
Expert 3 (3 images without annotated lesions*)							
Annotated images	23	19	20	12	5	3	28
Annotated lesions	330	117	126	32	13	10	31
Annotated pixels	1485222	42621	1250969	174901	182030	294764	916951
Expert 4 (1 images without annotated lesions*)							
Annotated images	23	23	19	16	1	5	28
Annotated lesions	295	370	108	43	5	14	28
Annotated pixels	2430756	37963	1696670	168865	59249	639378	1028284
Expert fusion (2 images without HA,MA,HE or SE)							
Lesioned images	22	23	18	10	-	-	-
Lesions	217	90	67	17	-	-	-
Lesioned pixels	1094318	6507	1375181	85748	-	-	-

*Annotated lesions comprise HA,MA,HE,SE,IRMA and NV.

Table A.3: Test set characteristics of the DIARETDB1 database comprising expert annotations and the DIARETDB1 ground truth (result of expert annotation fusion), where HA = haemorrhages, where MA = mircoaneurysms, HE = hard exudates, SE = soft exudates, IRMA = intra retinal microvascular abnormalities, NV = neovascularisation and OD = optic disc.

Test set (61 images)							
No.	HA	MA	HE	SE	IRMA	NV	OD
Expert 1 (7 images without annotated lesions*)							
Annotated images	21	52	17	10	8	1	3
Annotated lesions	280	285	77	22	20	1	3
Annotated pixels	1364825	81515	1476752	139698	148048	12197	64639
Expert 2 (23 images without annotated lesions*)							
Annotated images	23	30	23	16	3	1	58
Annotated lesions	184	286	95	42	5	1	58
Annotated pixels	2368297	2148457	2751463	318013	66525	21789	2305892
Expert 3 (23 images without annotated lesions*)							
Annotated images	24	27	26	14	2	1	60
Annotated lesions	301	142	134	39	5	1	62
Annotated pixels	1022537	35630	977363	150017	21670	6873	1947036
Expert 4 (19 images without annotated lesions*)							
Annotated images	26	35	26	16	-	1	59
Annotated lesions	368	500	104	66	-	1	59
Annotated pixels	1555852	44158	1928453	213691	-	13401	2265915
Expert fusion (20 images without HA,MA,HE and SE)							
Lesioned images	21	33	20	11	-	-	-
Lesions	187	116	75	18	-	-	-
Lesioned pixels	877586	8352	1237196	63902	-	-	-

*Annotated lesions comprise HA,MA,HE,SE,IRMA and NV.

A. Database characteristics

Table A.4: List of DIARETDB1 train set images.

diaretdb1_image001, diaretdb1_image003, diaretdb1_image004, diaretdb1_image009,
diaretdb1_image010, diaretdb1_image011, diaretdb1_image013, diaretdb1_image014,
diaretdb1_image020, diaretdb1_image022, diaretdb1_image024, diaretdb1_image025,
diaretdb1_image026, diaretdb1_image027, diaretdb1_image028, diaretdb1_image029,
diaretdb1_image035, diaretdb1_image036, diaretdb1_image038, diaretdb1_image042
diaretdb1_image053, diaretdb1_image055, diaretdb1_image057, diaretdb1_image064
diaretdb1_image070, diaretdb1_image079, diaretdb1_image084, diaretdb1_image086.

Table A.5: List of DIARETDB1 test set images.

diaretdb1_image002, diaretdb1_image005, diaretdb1_image006, diaretdb1_image007,
diaretdb1_image008, diaretdb1_image012, diaretdb1_image015, diaretdb1_image016,
diaretdb1_image017, diaretdb1_image018, diaretdb1_image019, diaretdb1_image021,
diaretdb1_image023, diaretdb1_image030, diaretdb1_image031, diaretdb1_image032,
diaretdb1_image033, diaretdb1_image034, diaretdb1_image037, diaretdb1_image039,
diaretdb1_image040, diaretdb1_image041, diaretdb1_image044, diaretdb1_image043,
diaretdb1_image045, diaretdb1_image046, diaretdb1_image047, diaretdb1_image048,
diaretdb1_image049, diaretdb1_image050, diaretdb1_image051, diaretdb1_image052,
diaretdb1_image054, diaretdb1_image056, diaretdb1_image058, diaretdb1_image059,
diaretdb1_image060, diaretdb1_image061, diaretdb1_image062, diaretdb1_image063,
diaretdb1_image065, diaretdb1_image066, diaretdb1_image067, diaretdb1_image068,
diaretdb1_image069, diaretdb1_image071, diaretdb1_image072, diaretdb1_image073,
diaretdb1_image074, diaretdb1_image075, diaretdb1_image076, diaretdb1_image077,
diaretdb1_image078, diaretdb1_image080, diaretdb1_image081, diaretdb1_image082,
diaretdb1_image083, diaretdb1_image085, diaretdb1_image087, diaretdb1_image088,
diaretdb1_image089.

Table A.6: Train and test characteristics of the BRISTOLDB database.

	Train set (30 images)	Test set (77 images)
Hard exudate images	30	60
Hard exudates	1393	2060
Hard exudate pixels	45231	55178

Image-based evaluation results (ROC curves)

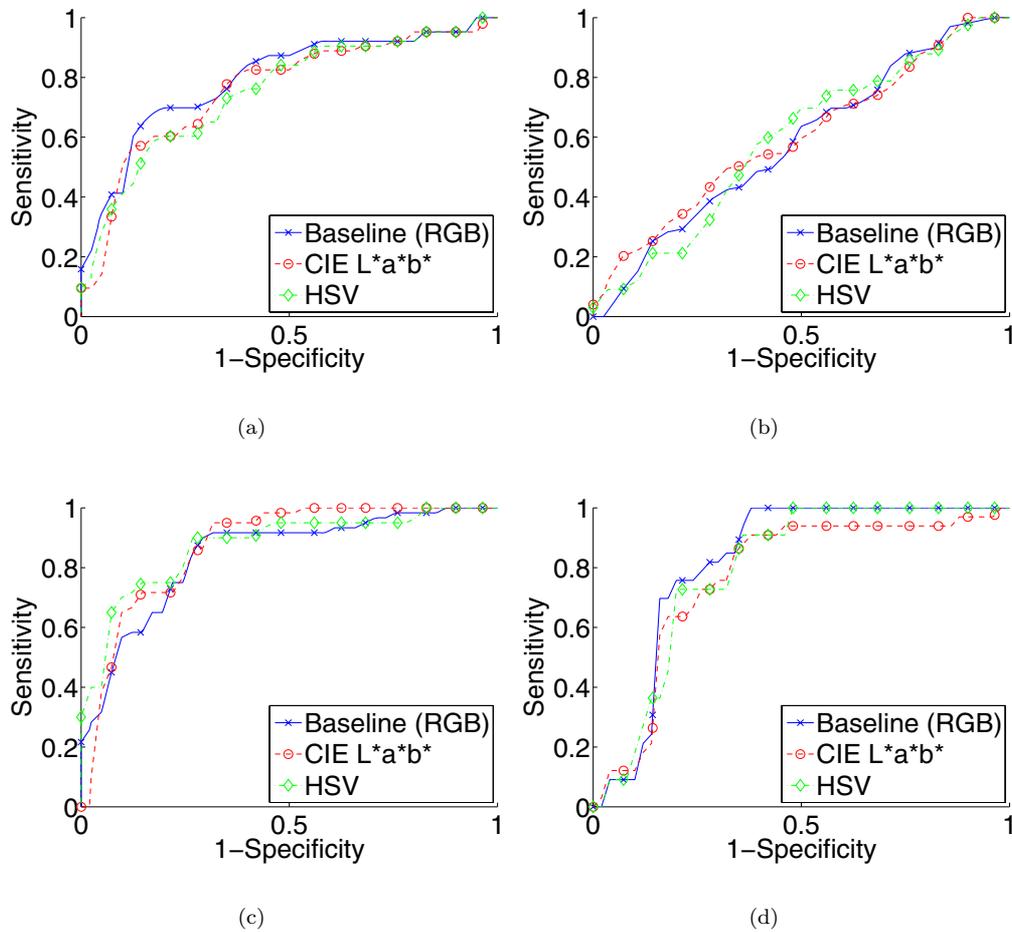


Figure B.1: Mean ROC curves (image-based evaluation) for the the baseline algorithm using different colour spaces: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

B. Image-based evaluation results (ROC curves)

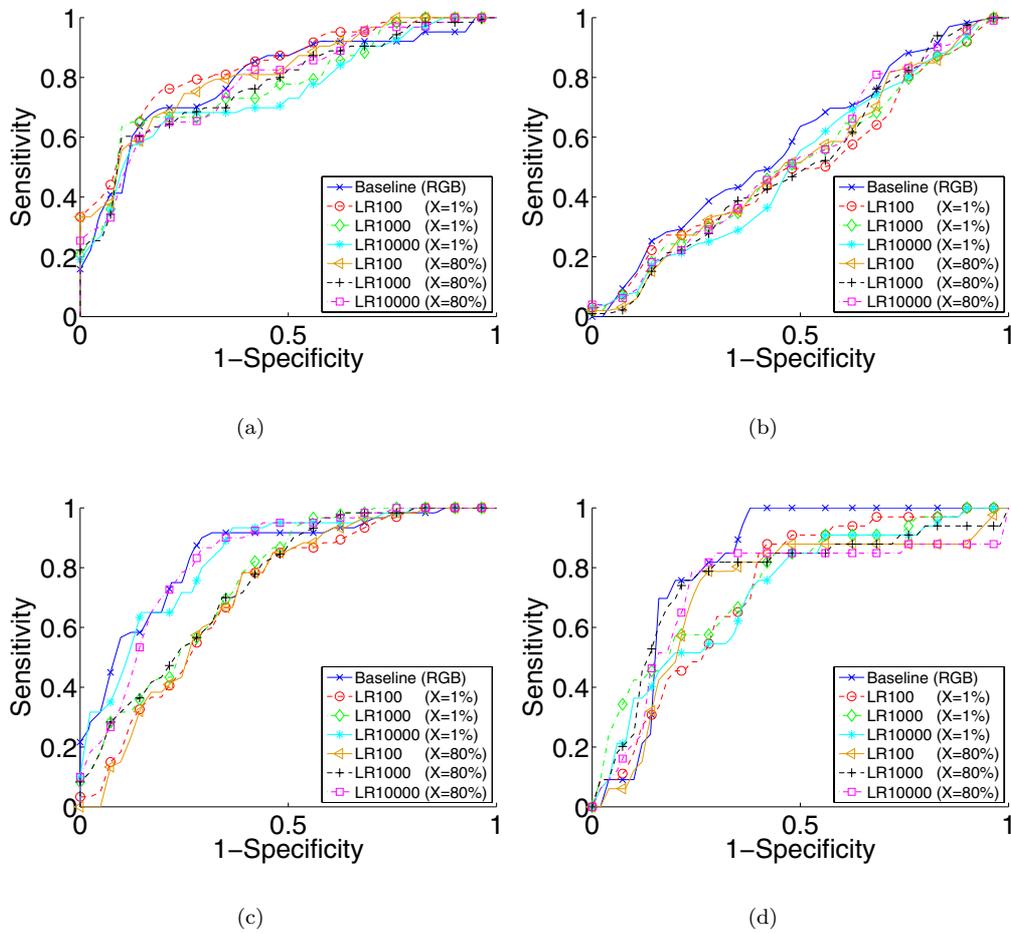


Figure B.2: Mean ROC curves (image-based evaluation with summax rule parameteres: $X=1\%$ and $X=80\%$) for the baseline algorithm using the background class information: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

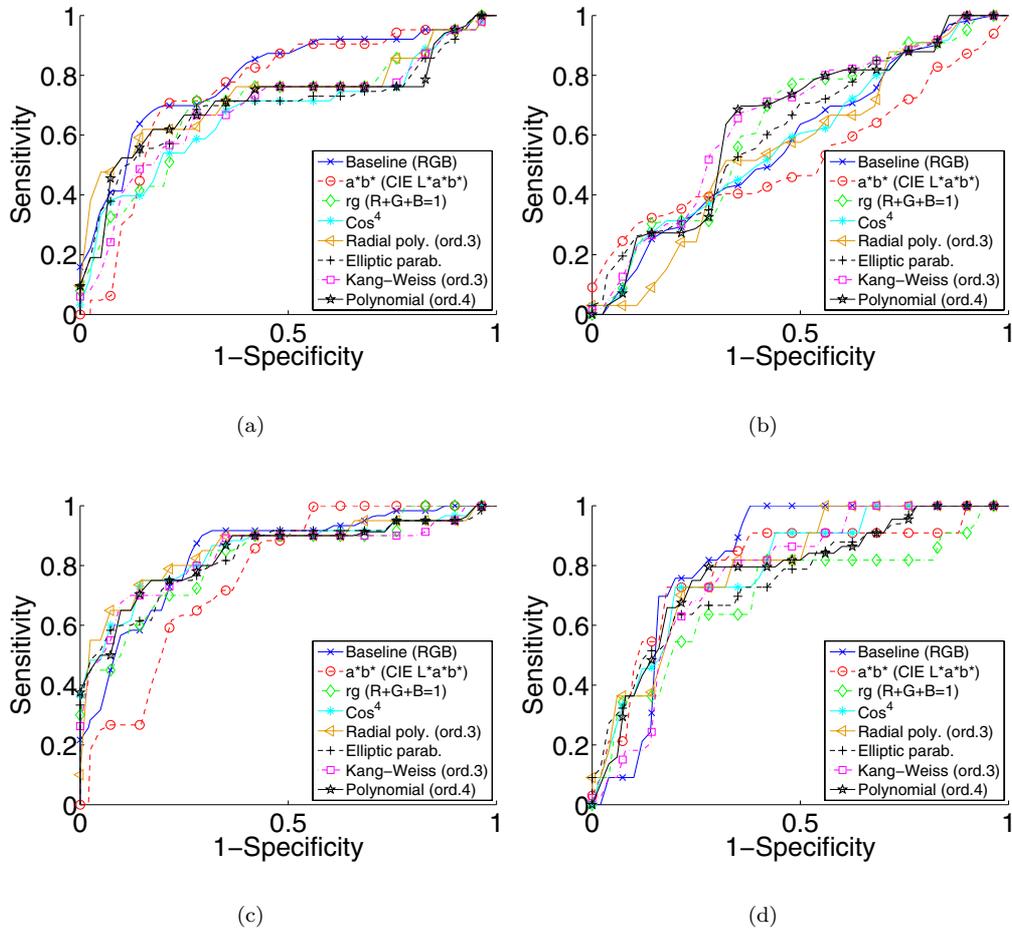


Figure B.3: Mean ROC curves (image-based evaluation) for the baseline algorithm using illuminance correction: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

B. Image-based evaluation results (ROC curves)

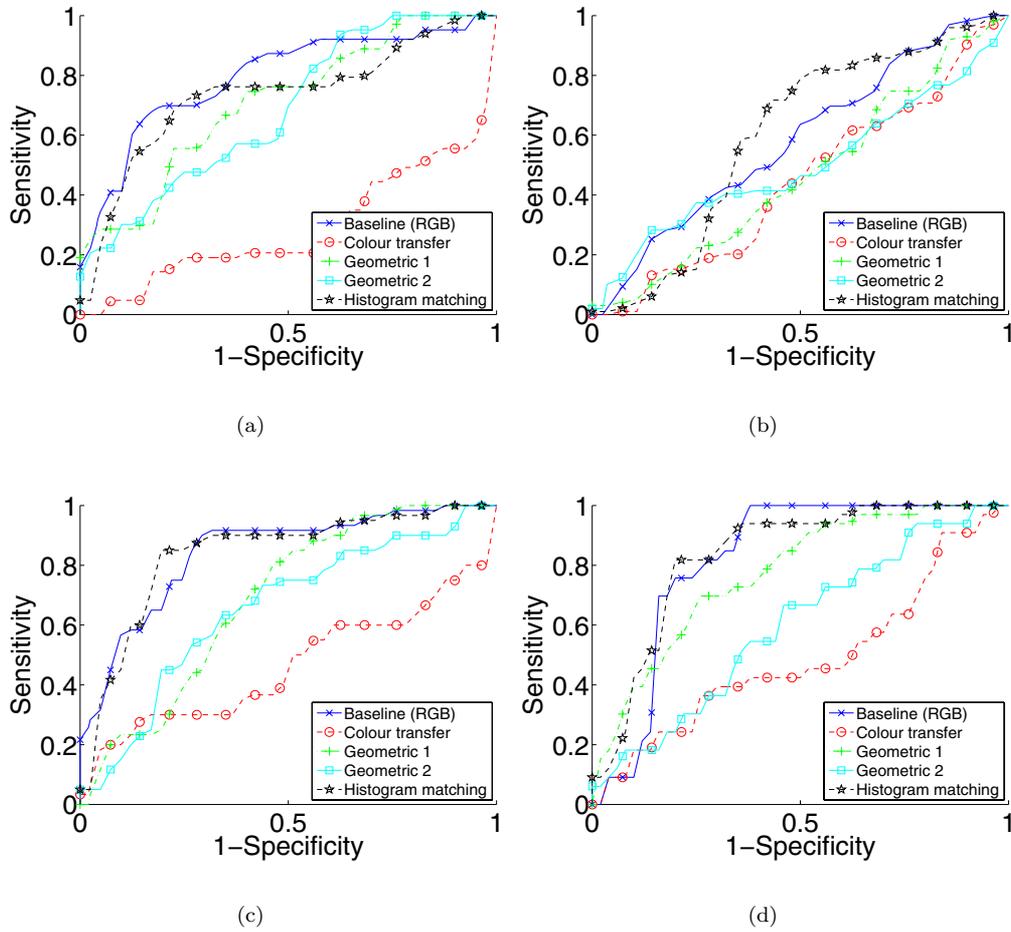


Figure B.4: Mean ROC curves (image-based evaluation) for the baseline algorithm using colour correction: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

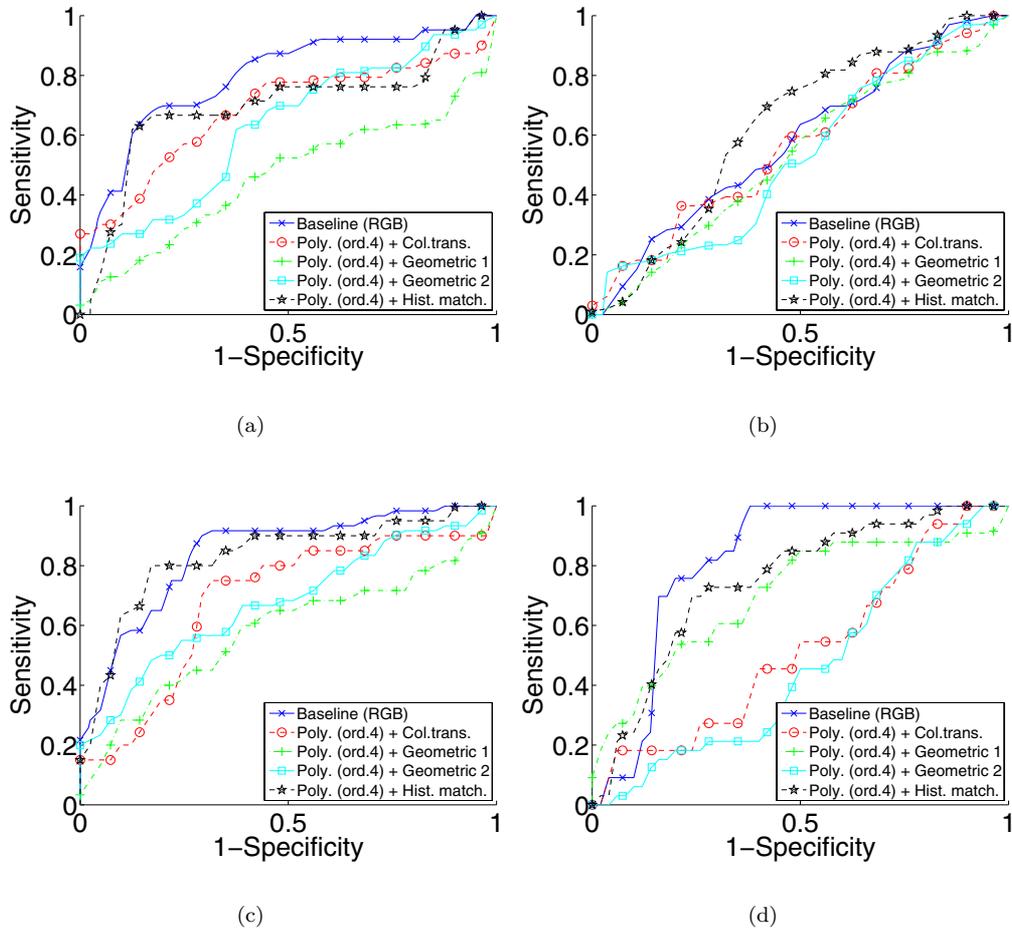


Figure B.5: Mean ROC curves (image-based evaluation) for the baseline algorithm using illumination and colour correction: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

Pixel-based evaluation results (ROC curves)

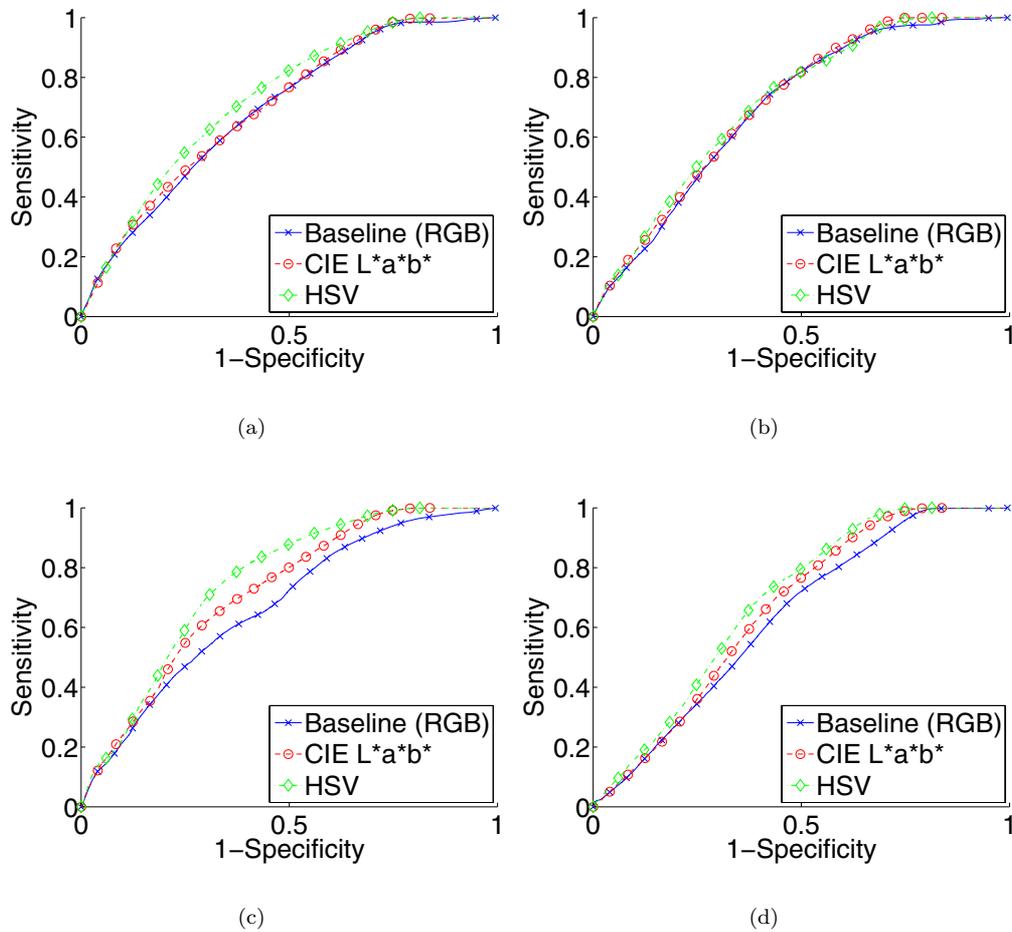


Figure C.1: Mean ROC curves (pixel-based evaluation) for the baseline algorithm using different colour spaces: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

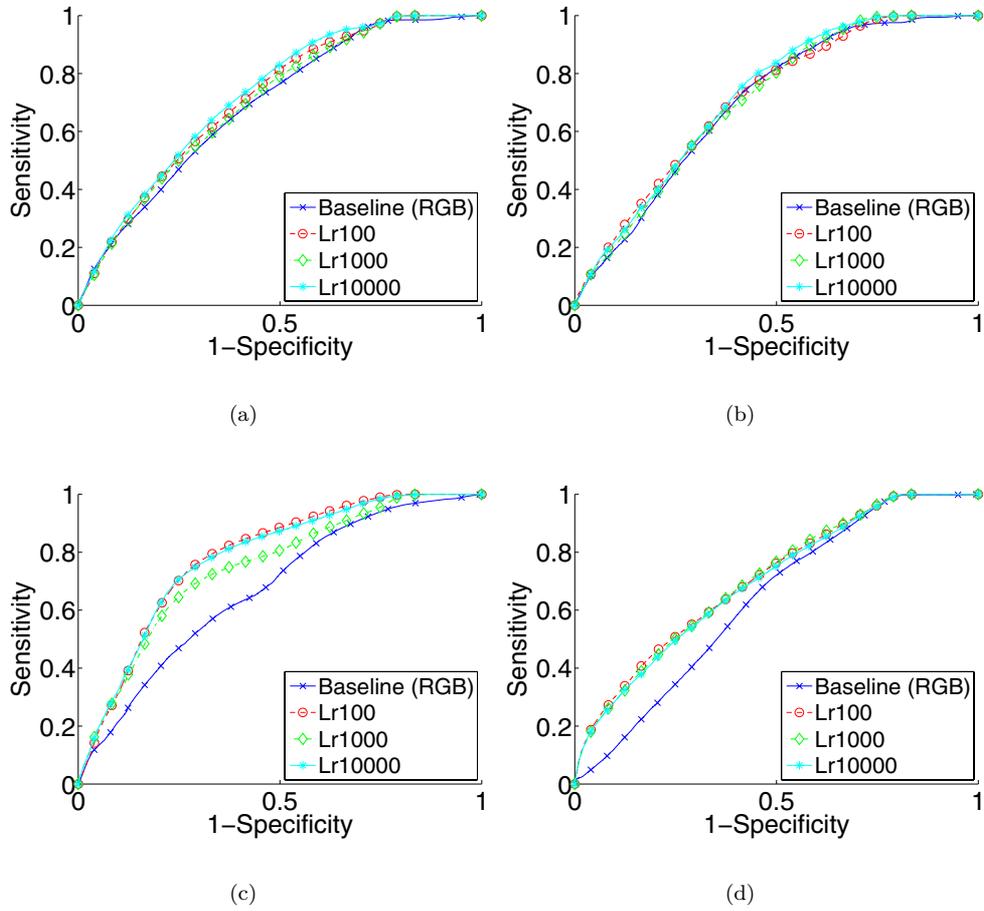


Figure C.2: Mean ROC curves (pixel-based evaluation) for the baseline algorithm using the background class information: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

C. Pixel-based evaluation results (ROC curves)

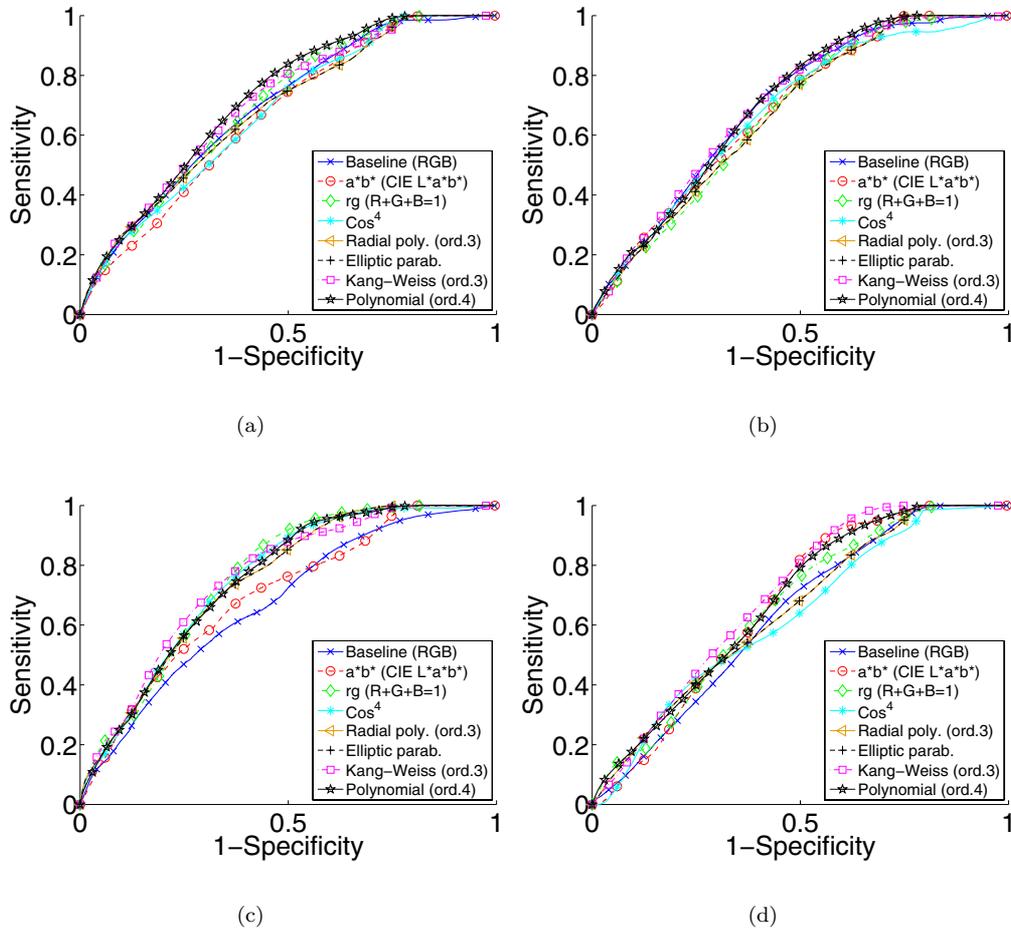


Figure C.3: Mean ROC curves (pixel-based evaluation) for the baseline algorithm using illumination correction: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

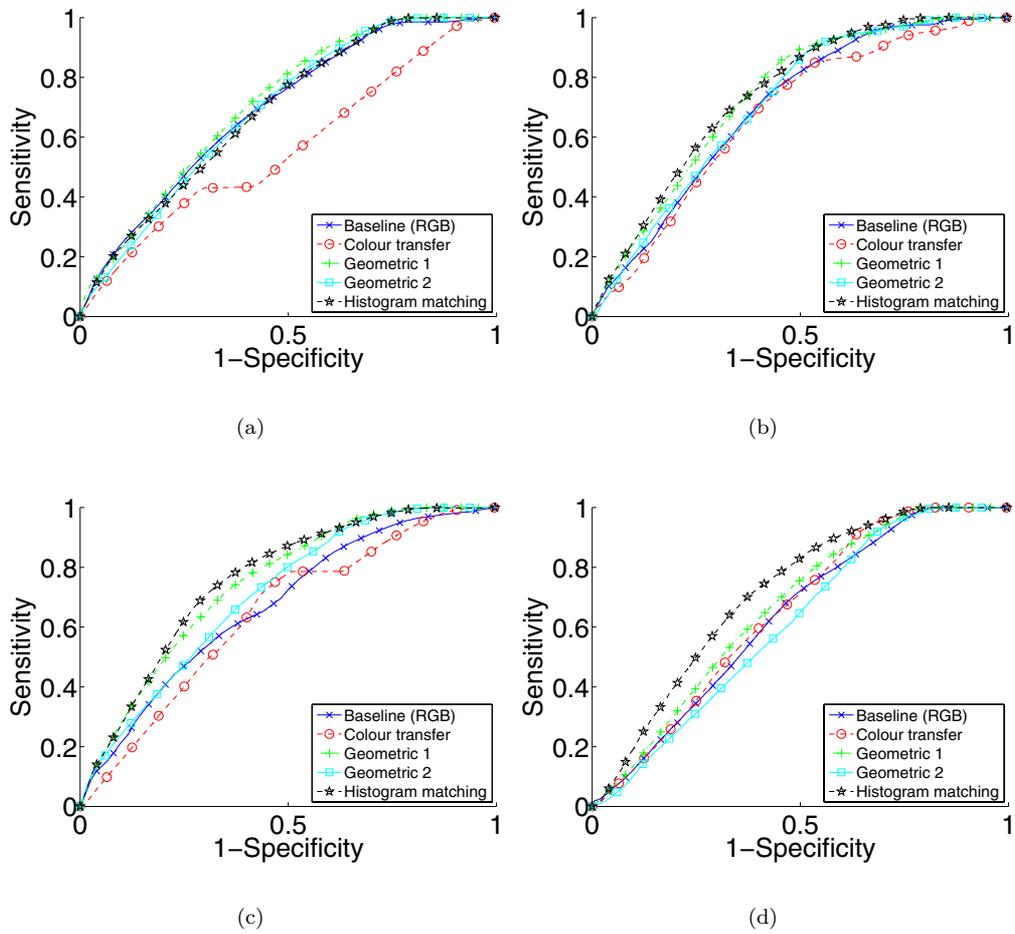


Figure C.4: Mean ROC curves (pixel-based evaluation) for the baseline algorithm using colour correction: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

C. Pixel-based evaluation results (ROC curves)

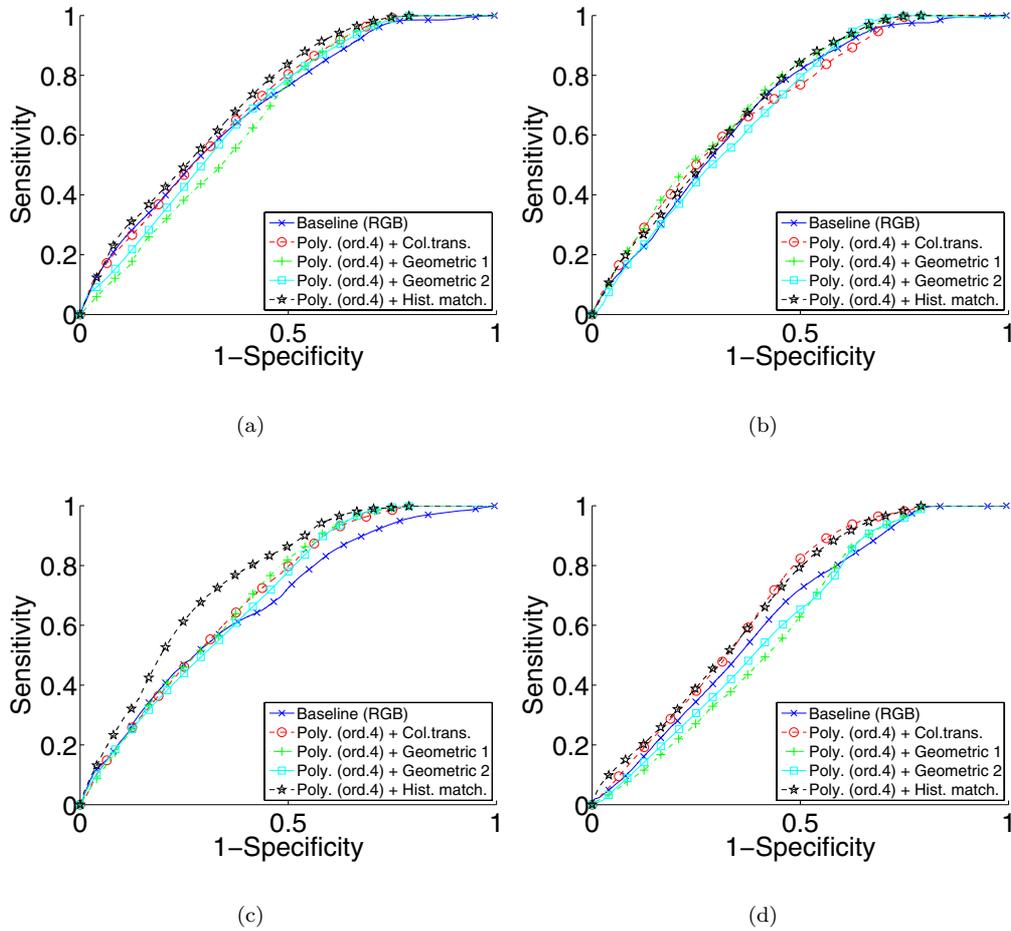


Figure C.5: Mean ROC curves (pixel-based evaluation) for the baseline algorithm using illumination and colour correction: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

Training error of baseline algorithm

Table D.1: Mean EER results (image-based evaluation) for the baseline algorithm using RGB colour space. Due to the stochastic nature of GMM-FJ initialisation, training is conducted three times for the train set. Testing and evaluation for both sets are conducted using the same trained classifiers. The image-based results for train set show clearly worse results for soft exudates which may imply that the detection problem is more difficult when diabetic retinopathy is strongly present. On the other, the number of lesioned images (Table A.2) depending on the lesion type is either too high or too low and few unlucky missclassifications may affect the performance dramatically.

	HA	MA	HE	SE	Avg.
Test set (RGB)	0.29	0.48	0.24	0.25	0.31
Train set	0.25	0.40	0.30	0.38	0.33

Table D.2: Mean EER results (pixel-based evaluation) for the baseline algorithm using RGB colour space. Due to the stochastic nature of GMM-FJ initialisation, training is conducted three times for the train set. Testing and evaluation for both sets are then conducted using the same trained classifiers.

	<i>HA</i>	MA	HE	SE	Avg.
Test set (RGB)	0.37	0.36	0.39	0.41	0.38
Train set (RGB)	0.38	0.33	0.34	0.27	0.33

D. Training error of baseline algorithm

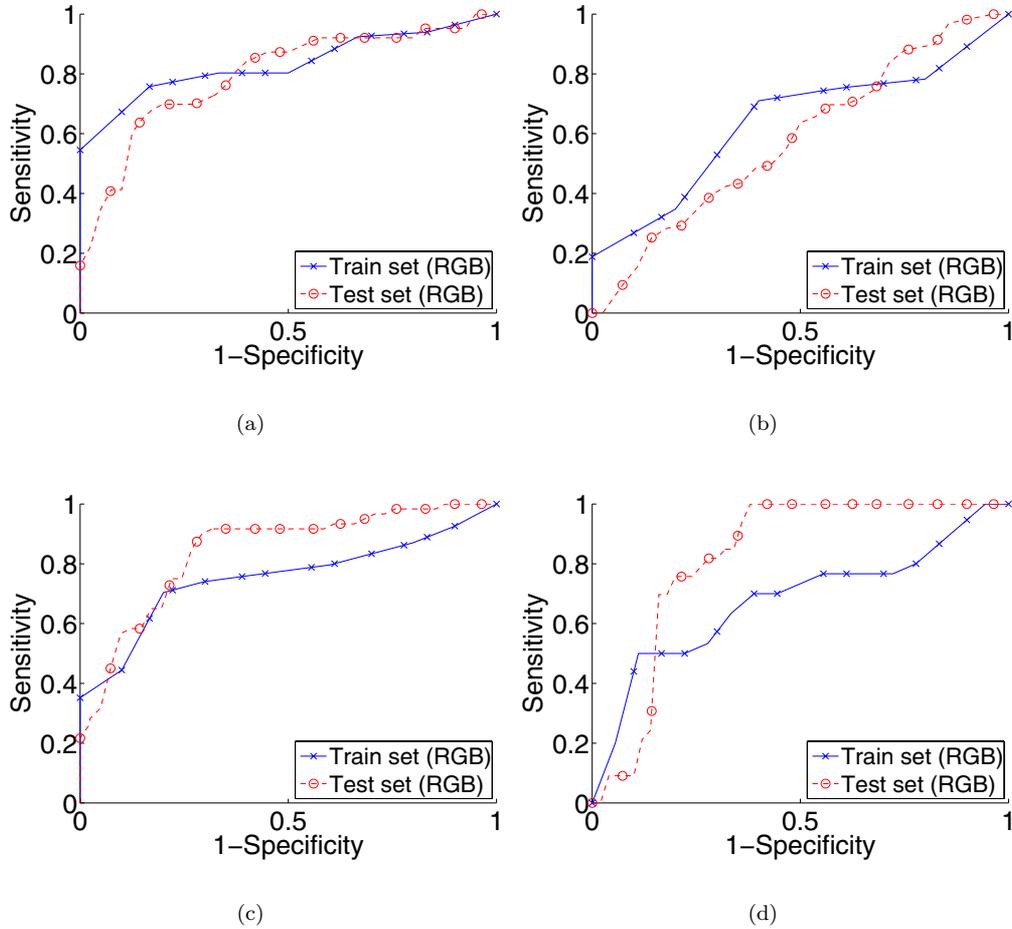


Figure D.1: Mean ROC curves (image-based evaluation) showing the training error for the baseline algorithm using RGB colour information. The image-based results for train set show clearly worse results for soft exudates which may imply that the detection problem is more difficult when diabetic retinopathy is strongly present. On the other, the number of lesioned images depending on the lesion type is either too high or too low and few unlucky misclassifications may affect the performance dramatically. (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

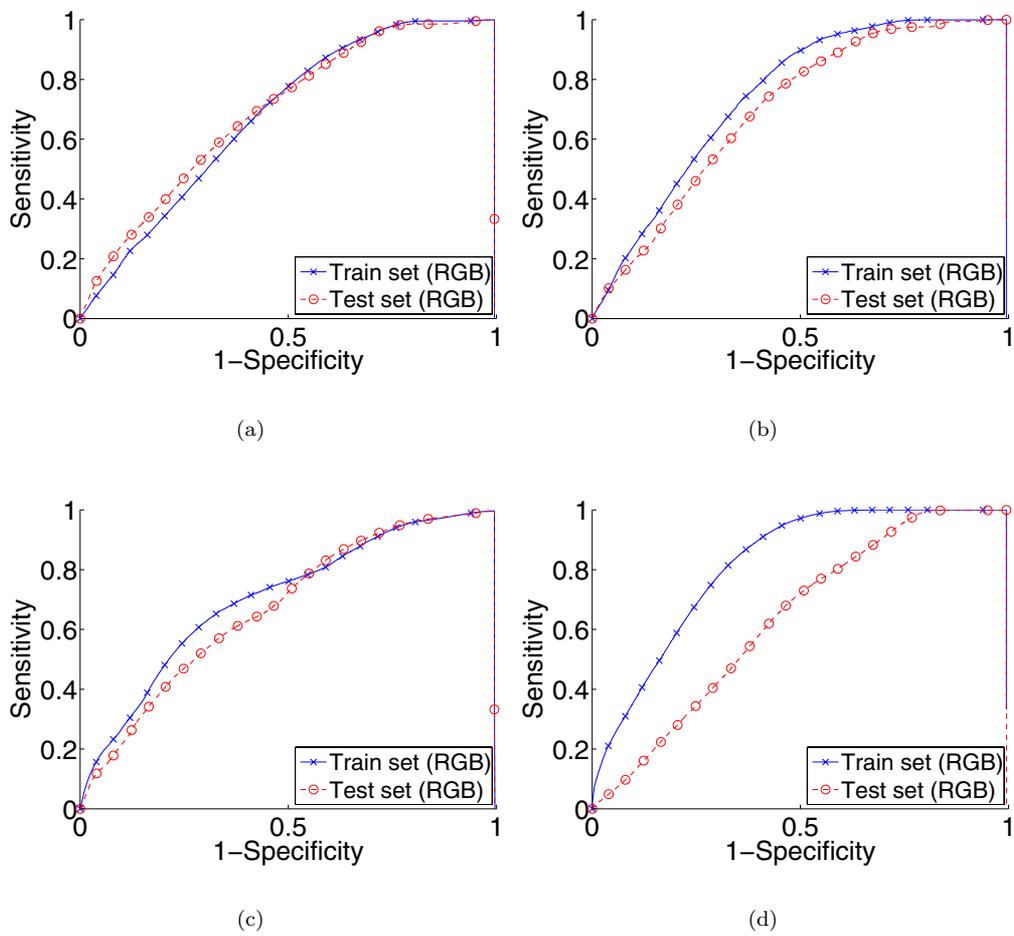


Figure D.2: Mean ROC curves (pixel-based evaluation) showing the training error for the baseline algorithm using RGB colour information.: (a) haemorrhages; (b) microaneurysms; (c) hard exudates; (d) soft exudates.

Example images of DIARETDB1

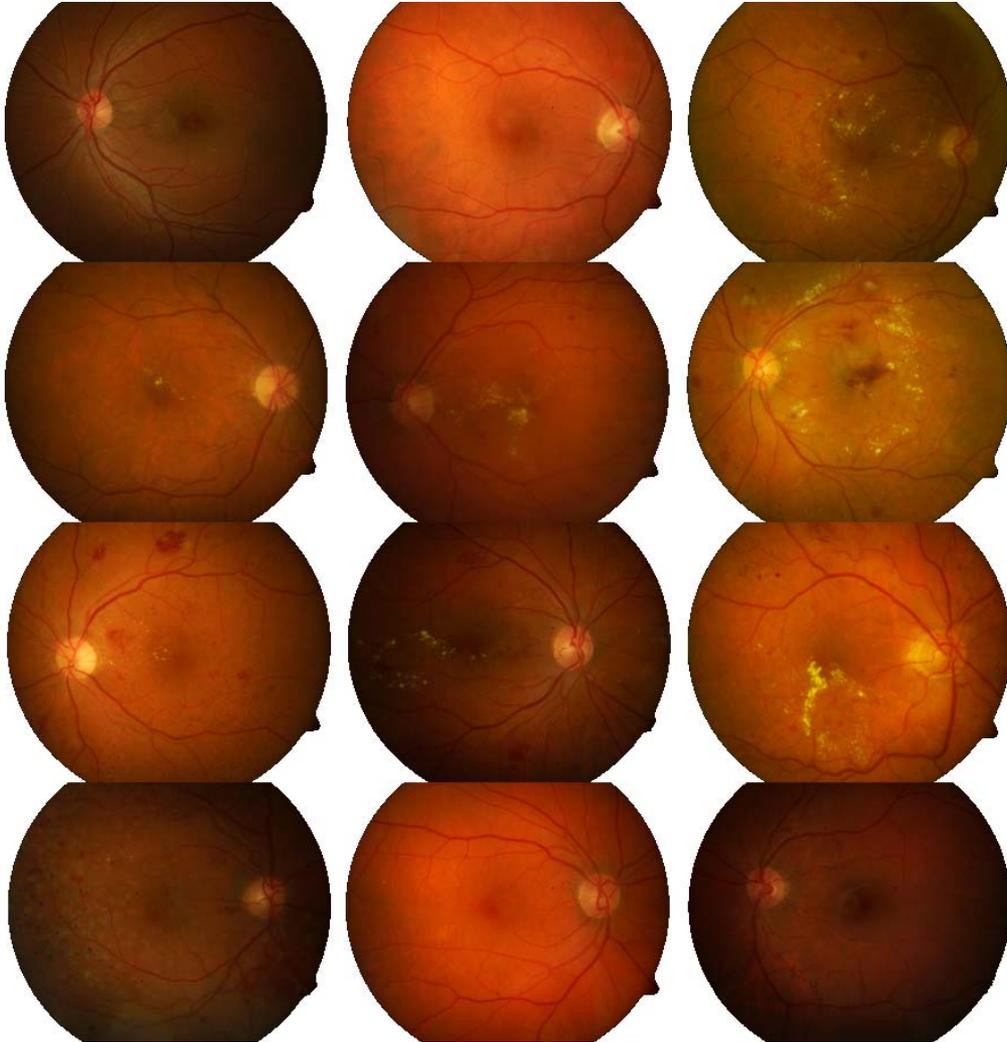


Figure E.1: Example images of public diabetic retinopathy database (DIARETDB1).

Illuminance corrected example images

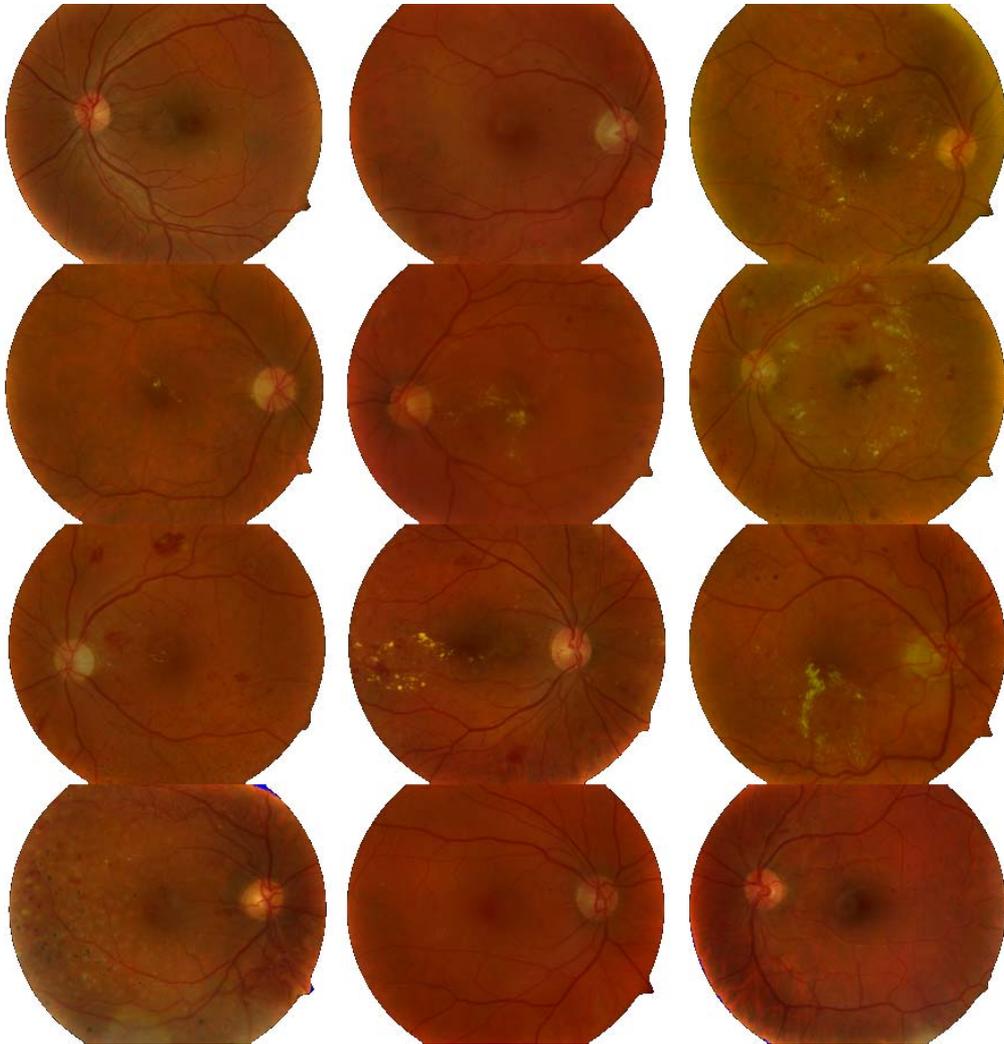


Figure F.1: Illuminance corrected images of public diabetic retinopathy database (DIARETDB1). The images are corrected using 4th order bivariate polynomial. Compare to the original images in Fig. E.1.

Colour corrected example images

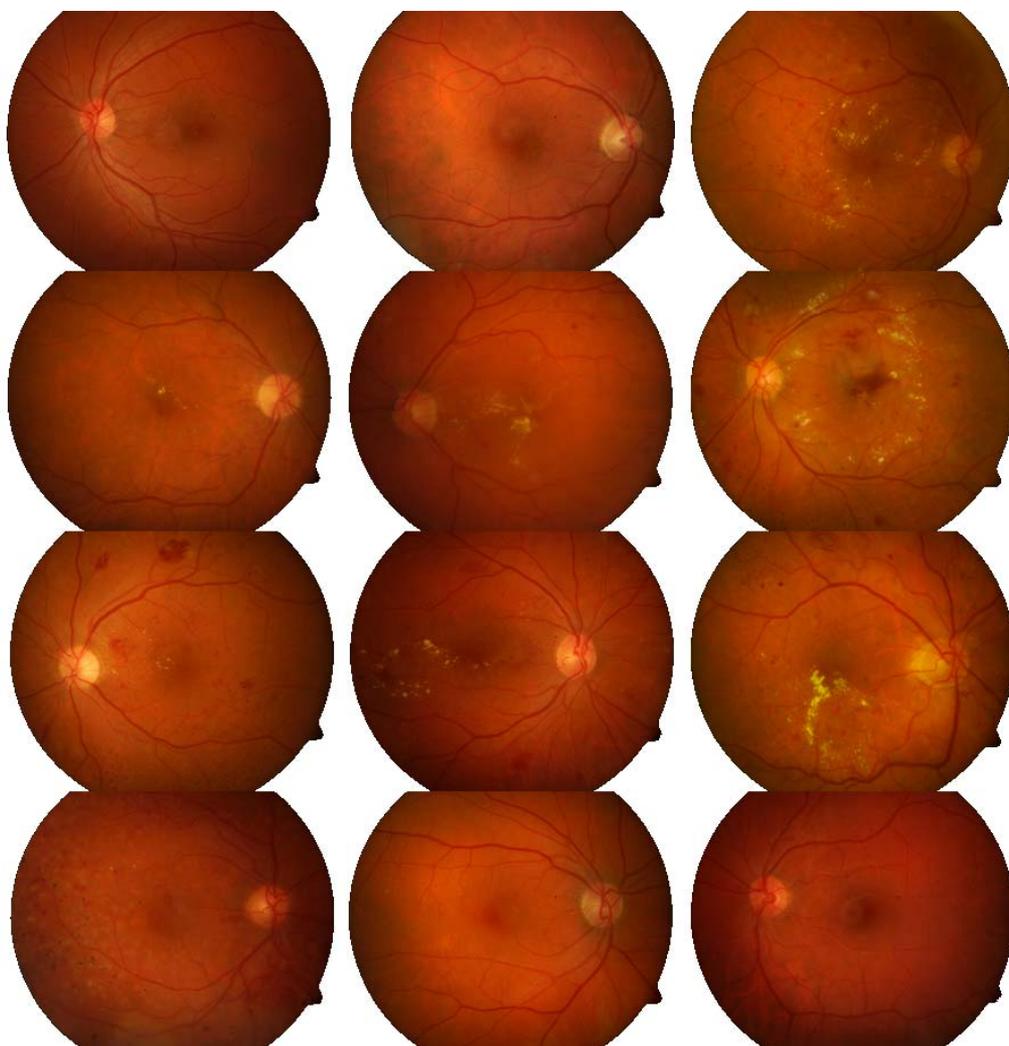


Figure G.1: Colour corrected images of public diabetic retinopathy database (DIARETDB1). The images are corrected using histogram matching. Compare to the original images in Fig. E.1.

Illuminance and colour corrected example images

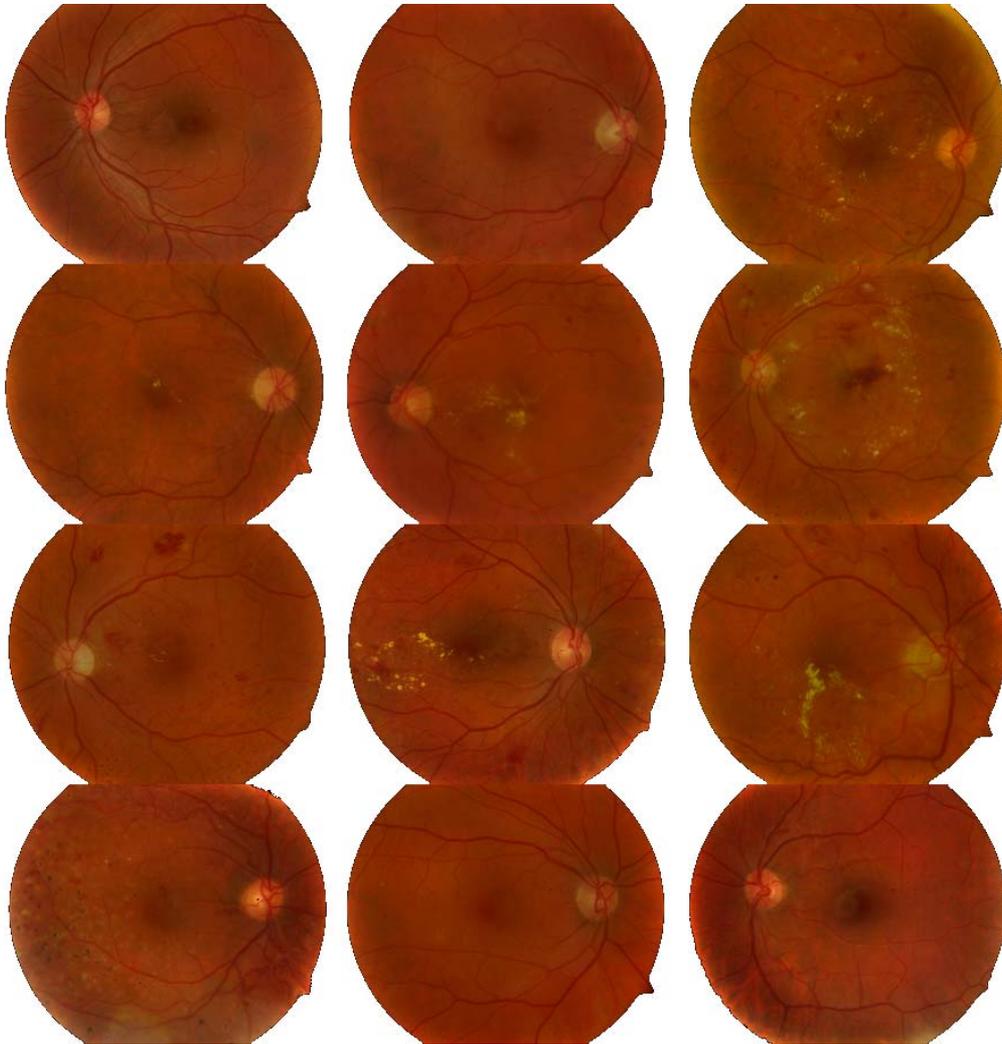


Figure H.1: Illuminance and colour corrected images of public diabetic retinopathy database (DIARETDB1). The images are corrected using 4th order bivariate polynomial and histogram matching. Compare to the original images in Fig. E.1

ACTA UNIVERSITATIS LAPPEENRANTAENSIS

370. HENTTONEN, KAISA. The effects of social networks on work-team effectiveness. 2009. Diss.
371. LASSILA, JUKKA. Strategic development of electricity distribution networks – Concept and methods. 2009. Diss.
372. PAAKKUNAINEN, MAARET. Sampling in chemical analysis. 2009. Diss.
373. LISUNOV, KONSTANTIN. Magnetic and transport properties of II-V diluted magnetic semiconductors doped with manganese and nickel. 2009. Diss.
374. JUSSILA, HANNE. Concentrated winding multiphase permanent magnet machine design and electromagnetic properties – Case axial flux machine. 2009. Diss.
375. AUVINEN, HARRI. Inversion and assimilation methods with applications in geophysical remote sensing. 2009. Diss.
376. KINDSIGO, MERIT. Wet oxidation of recalcitrant lignin waters: Experimental and kinetic studies. 2009. Diss.
377. PESSI, PEKKA. Novel robot solutions for carrying out field joint welding and machining in the assembly of the vacuum vessel of ITER. 2009. Diss.
378. STRÖM, JUHA-PEKKA. Activated u/dt filtering for variable-speed AC drives. 2009. Diss.
379. NURMI, SIMO A. Computational and experimental investigation of the grooved roll in paper machine environment. 2009. Diss.
380. HÄKKINEN, ANTTI. The influence of crystallization conditions on the filtration characteristics of sulphathiazole suspensions. 2009. Diss.
381. SYRJÄ, PASI. Pienten osakeyhtiöiden verosuunnittelu – empiirinen tutkimus. 2010. Diss.
382. KERKKÄNEN, ANNASTIINA. Improving demand forecasting practices in the industrial context. 2010. Diss.
383. TAHVANAINEN, KAISA. Managing regulatory risks when outsourcing network-related services in the electricity distribution sector. 2010. Diss.
384. RITALA, PAAVO. Cooperative advantage – How firms create and appropriate value by collaborating with their competitors. 2010. Diss.
385. RAUVANTO, IRINA. The intrinsic mechanisms of softwood fiber damage in brown stock fiber line unit operations. 2010. Diss.
386. NAUMANEN, VILLE. Multilevel converter modulation: implementation and analysis. 2010. Diss.
387. IKÄVALKO, MARKKU. Contextuality in SME ownership – Studies on owner-managers' ownership behavior. 2010. Diss.
388. SALOJÄRVI, HANNA. Customer knowledge processing in key account management. 2010. Diss.
389. ITKONEN, TONI. Parallel-operating three-phase voltage source inverters – Circulating current modeling, analysis and mitigation. 2010. Diss.
390. EEROLA, TUOMAS. Computational visual quality of digitally printed images. 2010. Diss.

391. TIAINEN, RISTO. Utilization of a time domain simulator in the technical and economic analysis of a wind turbine electric drive train. 2010. Diss.
392. GRÖNMAN AKI. Numerical modelling of small supersonic axial flow turbines. 2010. Diss.
393. KÄHKÖNEN, ANNI-KAISA. The role of power relations in strategic supply management – A value net approach. 2010. Diss.
394. VIROLAINEN, ILKKA. Johdon coaching: Rajanvetoja, taustateorioita ja prosesseja. 2010. Diss.
395. HONG, JIANZHONG. Cultural aspects of university-industry knowledge interaction. 2010. Diss.
396. AARNIOVUORI, LASSI. Induction motor drive energy efficiency – Simulation and analysis. 2010. Diss.
397. SALMINEN, KRISTIAN. The effects of some furnish and paper structure related factors on wet web tensile and relaxation characteristics. 2010. Diss.
398. WANDERA, CATHERINE. Performance of high power fiber laser cutting of thick-section steel and medium-section aluminium. 2010. Diss.
399. ALATALO, HANNU. Supersaturation-controlled crystallization. 2010. Diss.
400. RUNGI, MAIT. Management of interdependency in project portfolio management. 2010. Diss.
401. PITKÄNEN, HEIKKI. First principles modeling of metallic alloys and alloy surfaces. 2010. Diss.
402. VAHTERISTO, KARI. Kinetic modeling of mechanisms of industrially important organic reactions in gas and liquid phase. 2010. Diss.
403. LAAKKONEN, TOMMI. Distributed control architecture of power electronics building-block-based frequency converters. 2010. Diss.
404. PELTONIEMI, PASI. Phase voltage control and filtering in a converter-fed single-phase customer-end system of the LVDC distribution network. 2010. Diss.
405. TANSKANEN, ANNA. Analysis of electricity distribution network operation business models and capitalization of control room functions with DMS. 2010. Diss.
406. PIIRAINEN, KALLE A. IDEAS for strategic technology management: Design of an electronically mediated scenario process. 2010. Diss.
407. JOKINEN, MARKKU. Centralized motion control of a linear tooth belt drive: Analysis of the performance and limitations. 2010. Diss.
408. KÄMÄRI, VESA. Kumppanuusohjelman strateginen johtaminen – Monitapaustutkimus puolustushallinnossa. 2010. Diss.
409. KARJALAINEN, AHTI. Online ultrasound measurements of membrane compaction. 2010. Diss.
410. LOHTANDER, MIKA. On the development of object functions and restrictions for shapes made with a turret punch press. 2010. Diss.
411. SIHVO, VILLE. Insulated system in an integrated motor compressor. 2010. Diss.
412. SADOVNIKOV, ALBERT. Computational evaluation of print unevenness according to human vision. 2010. Diss.
413. SJÖGREN, HELENA. Osingonjakopäätökset pienissä osakeyhtiöissä. Empiirinen tutkimus osakeyhtiölain varojenjakosäännösten toteutumisesta. 2010. Diss.

